

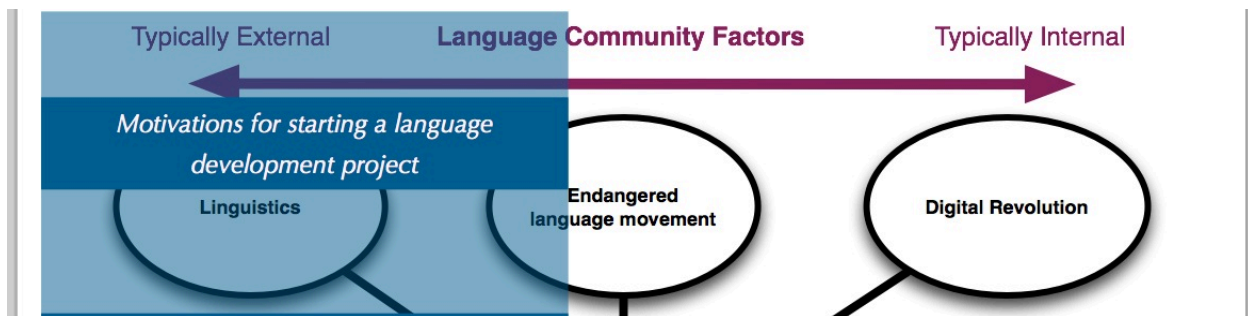
Pots of Knowledge based on diagram

Issues of desire

I outline some of the most salient issues in the text input task. (See full size diagrams at end of document.)

Motivations

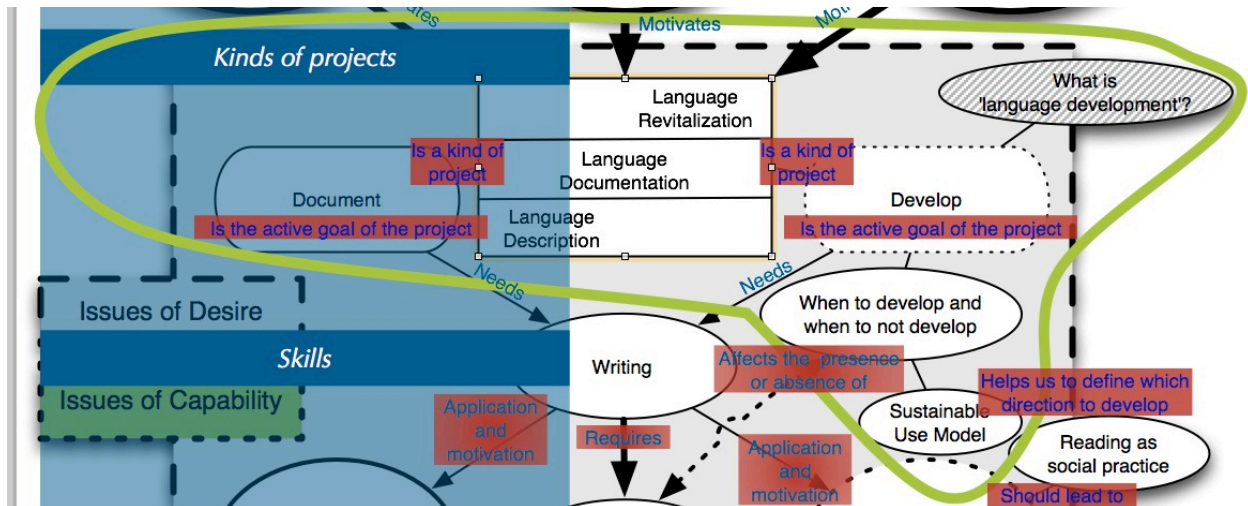
I say that there are three general motivations or “forces” which drive language description or language development projects. These three forces are *linguistic research agendas*, the *endangered language movement* and the *digital revolution*.



Projects

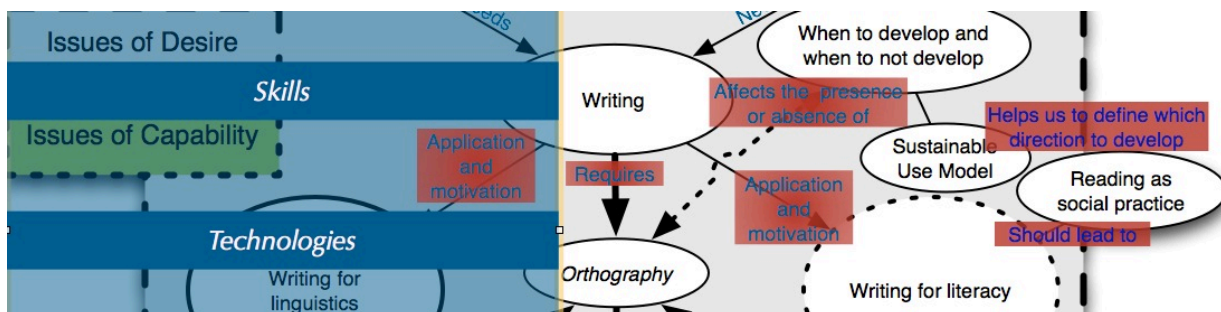
I say there are different kinds of projects which are conducted in and with communities. These include *Language Description*, *Language Documentation*, and *Language Revitalization* projects. I then further say that there are two basic tenants to these projects: **describe** the language (in a documented form with linguistic analysis), and to take steps to **develop** the language. Using the SUM model we can say that language development might be helping a speaker population move up or down a language vitality scale. In the process of planning a language communities' future, development goal need to be set. One of these goals is to clearly state whether or not to develop an orthography.

Both main activities - documenting and developing - require writing as a functional technology. These projects are the situations in which end users of keyboard technologies to realize their need for the technology.



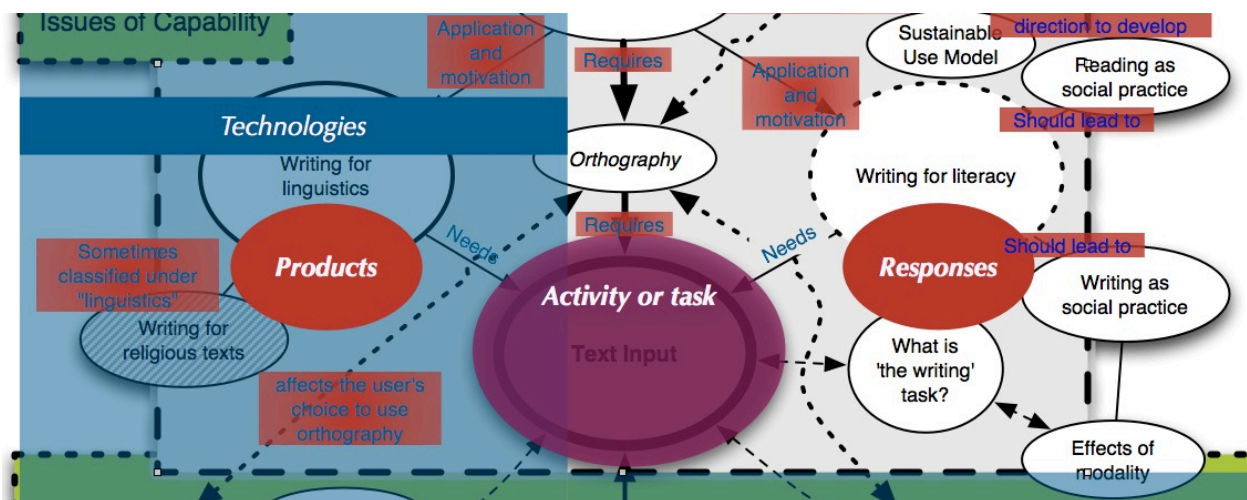
Skills

I say that writing can be both a skill or a technology. The technology of writing is required if we are going to develop an Orthography, or writing based products like linguistic texts or religious texts. Writing is also required as a technology if literacy in the language is going to be expected. So, I see writing as something more fundamental than an orthography, and something required by social literacy and linguistic descriptions.



Social practice

I see writing as separate social practice from 'reading as a social practice'. I see writing as a technology as fundamental to both reading and writing as social practices, but I also see 'writing as a social practice' as being more important than 'reading as a social practice' (in terms of development goals). When people are writing they are reading. When people are reading they might not be writing. For a language to maintain its vitality, it needs new knowledge to be expressed in the language. New knowledge is not just reading it means expressing too.



There are difference between writing for social practice and writing for projects. One key difference, is that writing for social practice is writing with an intent to elicit a *response* whereas writing in a project writes to create a *product*. Both of these goals require the activity of text input or the task of text input.

When considering the design of new text input methods for languages with small speaker populations it is important to take into account the design target. That is, are the design specifications for product production or to encourage writing as a social practice. If the target is writing as social practice, does this kind of design call for ‘just

good enough' solutions or solutions which 'out perform' solutions in competing languages?

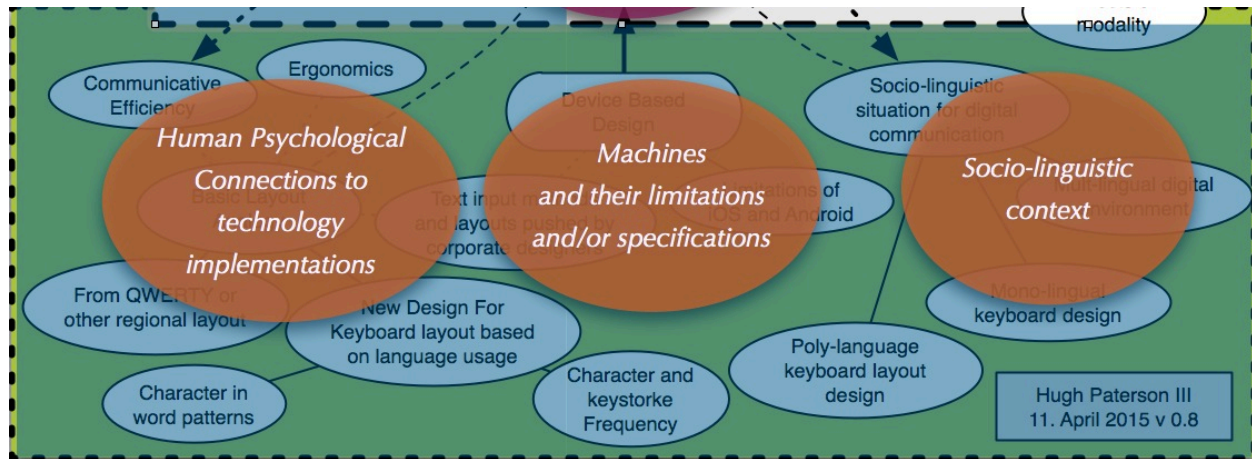
The activity or task of text Input

The task or activity of text input takes place in different mediums or modalities. For instance responses are often sought through SMS messages (often on small screens and phones), whereas books or journal articles are designed to stand on their own (which are almost exclusively typed via full size keyboards on laptops and computers).

Text input solutions with keyboards require keyboard layouts. Therefore keyboard layouts become a foundation of all the previously mentioned item. They are used by language projects, they are used in the social practice of reading and writing, they are used in the production of content from language projects.

Issues of capability

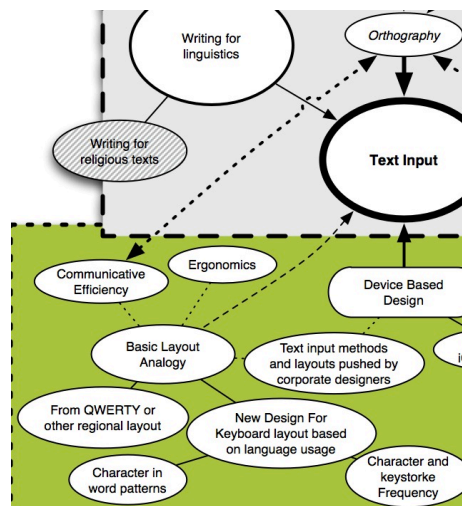
I outline several groups of issues in keyboard layouts. I group these issues in to three main groups as follows: *Human psychological connections to technology implementations, Machines and their limitations and/or specifications, and Socio-linguistic context*. These issues affect the capability of the language users to use a given text input solution with their language.



Human Psychological Connections to technology implementations

There are several factors or issues which I lump under this section. These include:

Communicative efficiency, Basic (keyboard) layout analogy, Influences from QWERTY or other regional layouts, Ergonomics, Character in word patterns, Character and keystroke frequency, New design for keyboard layout based on language usage, Text



input methods and layouts pushed by corporate designers.

I see these factors working together in the ensuing described ways.

Communicative Efficiency

Every language in both oral and written form has a way to discuss frequently expressed ideas in addition to less frequently discussed, or new ideas. Across

languages the connotations and quantity of vocabulary needed to communicate the same (or equivalent) ideas in different languages varies. For instance, the New Testament book of James, when expressed in written form in four different language takes on different numbers of characters and different numbers of words. Consider the following data from four languages:

Me'phaa

Character Count: 16618

Word Count: 2856

Sochiapam Chinantec

Character Count: 20416

Word Count: 4506

Spanish

Character Count: 9611

Word Count: 2165

English

Character Count: 10432

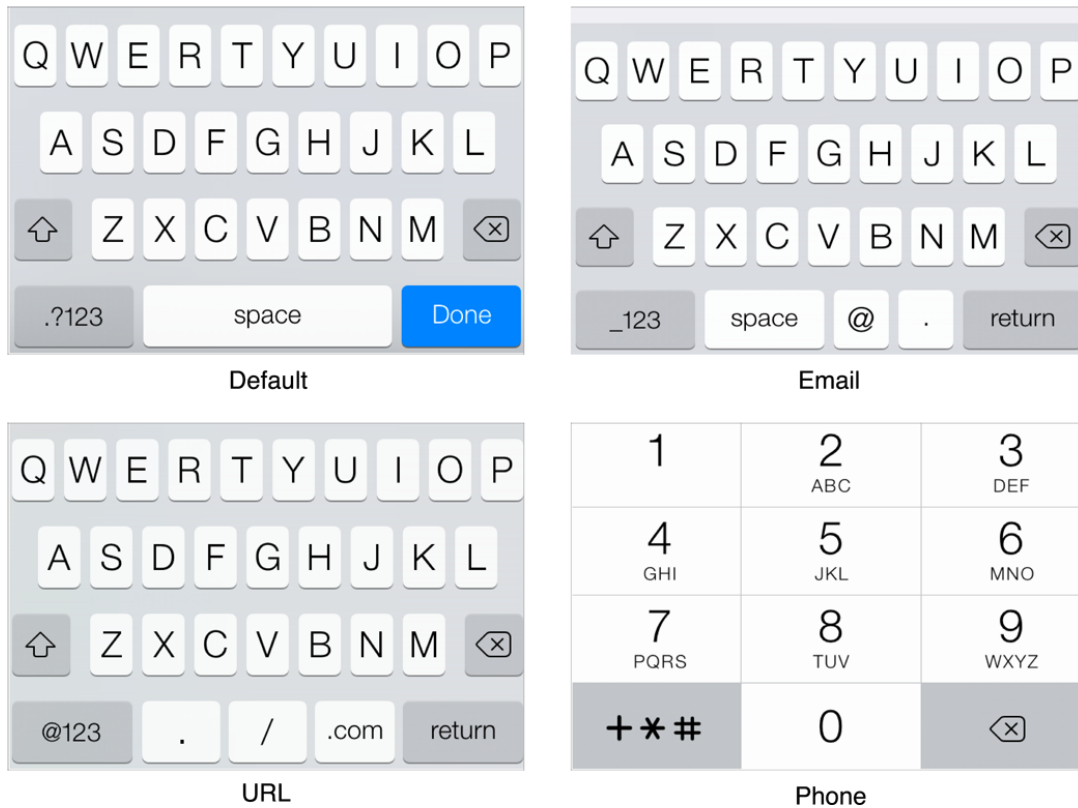
Word Count: 2575

All other things being equal, Spanish can very easily be considered the most efficient of the four languages to communicate in because it has the fewest words and the fewest characters. This means that users of this language would theoretically expend less effort to type this text and communicate the same ideas in Spanish when compared with the other options.

Communicative efficiency is necessarily dependent on a particular orthography. However, the orthography is not the only factor contributing to communicative efficiency. When determining communicative efficiency a speaker may choose between issues of affinity as expressed in language choice, or issues of practicality as dictated by technological medium. In terms of the text input task, the basic keyboard layout analogy often drives the perception of communicative efficiency in a given language.

Basic keyboard layout analogy

What I am calling a basic keyboard layout is an abstraction of a keyboard layout. For instance the keyboard layout QWERTY is slightly different on some cell phones based



on the purpose of the screen.¹ While most of the characters stay consistent, some functional buttons change positions. I also call this an analogy because the functional layout is close to a real physical layout. With touchscreen devices there is no physical constraint stipulating for the order of characters to be on the screen the same way as they appear on physical devices. Rather I posit that there is a psychological connection between the user and a mental construct (or keyboard arrangement). This construct may be based on learning or interactions done in a physical reality.

Sometimes the basic keyboard analogy for a given language can be influence by a globally dominant keyboard layout. For instance in the English speaking world

¹ Image from: Apple iOS Developer library. Managing the Keyboard. At: <https://developer.apple.com/library/ios/documentation/StringsTextFonts/Conceptual/TextAndWebiPhoneOS/KeyboardManagement/KeyboardManagement.html>

Also consider images and actions discussed here: <http://mashable.com/2010/08/19/iphone-keyboard-shortcuts-tips-tricks/>

QWERTY is a common layout, while in the French speaking world AZERTY is common.

This is to say that from region to region and from language to language the basic keyboard analogy which users have in their mental representation can be different.

Ergonomics

The basic keyboard analogy when applied to a physical form creates body movements in users as they utilize the keyboard layout. Previous studies have suggested that some more of the typing action should occur with stronger fingers. Ergonomics, based on a given device should then be a contributing factor to new keyboard layout designs.

Character in word patterns

Characters which occur in word patterns are dependent directly on the phonotactics of the language as they are realized in an orthography. The order and frequency of these patterns should have an impact on keyboard layouts because they affect distribution of characters in a keyboard layout. That is to create an ergonomically balanced layout frequent patterns need to be accounted for. One common way to account for these patterns are to consider them as n-grams and put each part of a n-gram on a different hand. Another approach would be to consider morphological input devices. That is, make certain keys dedicated to specific morphemes. This assumes of course that particular morphemes are responsible for the targeted patterns of characters in the language.

Character and keystroke frequency

Beside the character in word patterns there are more general patterns based on the frequency of characters in the language and the number of times a key is struck. For instance, a character in word pattern might consider a segmental morpheme [eng]

< ed > or trigraph such as < sch > in German. tonal markings of a language which marks tonal distinctions such as HL, LH, HLH, HML, etc. However, a keystroke frequency analysis might consider if dead keys are used to type these tonal marks, or additional characters. The following data from the Me'phaa translation of James may make the situation clearer.

Total uses of < a >: 2950

Total uses of < _ >: 1834 Keystrokes used: 3668 (this is across all vowels)

Total uses of < ´ >: 4477 Keystrokes used: 8954 (this is across all vowels)

Total uses of < a + _ >: 875 Keystrokes used: 1750

Total uses of < a + ´ >: 880 Keystrokes used: 1760

As we can see in this data < a > is some times used alone, or with one of two tonal diacritics. These diacritic uses form single orthographic characters but they are counted as digrams (n-gram where n = 2) when considering keystrokes (because dead keys are used). In the case of the Me'phaa text input system < a > is used more often with a diacritic than without. No text input system to my knowledge tracks the patterns of diacritics relative to each other. Such a system would be needed to implement logic based on tonal melodies (also know as tonal patterns).

New design for keyboard layout based on language usage

Language specific patterns of of keystrokes under ergonomic consideration should be the basis of keyboard layouts for new languages. In turn this new design layout should the basis for the *Basic (keyboard) layout analogy* specific to that language.

Corporate Design Influences

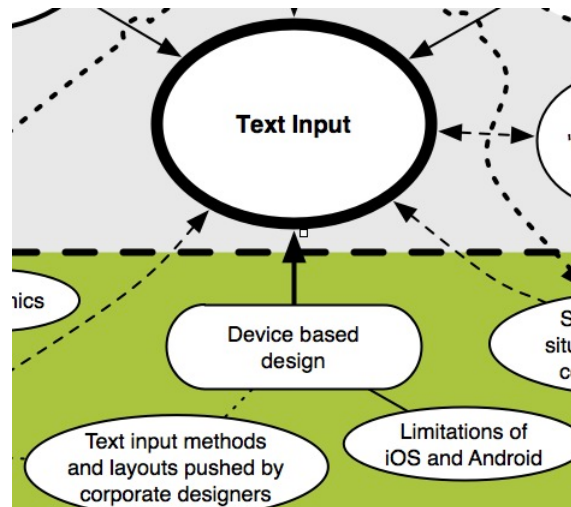
Text input methods and keyboard layouts are subject to the pressures of economic development. That is, manufacturers have the power to rapidly propagate text input solutions which benefit their economic interests. From a sociological perspective, this means that endusers are often at the whim of the options that manufacturers provide. Solutions for end users who are part of language communities with small populations are classified as 'edge cases'.

Let's be blunt about it. We may call these languages and their speakers by any of the following terms: minority languages, minority language speakers/users, ethnolinguistic minorities, languages with small speaker populations, under-developed languages, moribund languages, endangered languages, under-resourced languages, or low resourced languages. However, from a capitalistic and industrial complex point of view, these languages are really languages without financial viability. From the perspective of the inventor and capitalist seeking financial remuneration for solving problems for a wide range of customers, these languages hold very little if any return on investment.

It is partially for these financial reasons that text input solutions do not already exist. But it is also partially for these reasons that small languages are served text input solutions which are 'sub-optimal'.

Machines and their limitations and/or specifications

In this section I add to the issues started under Corporate Design Influences. But now I shift focus from input processes to the keyboard layout to input processes for the hardware. Well to the extent that hardware is the holistic embodiment of the hardware and software solution.



Limitations of iOS and Android

As previously discussed the input options pushed by corporate manufacturers play a role in what options are available to end users. However, devices (desktops, laptops, cell phones, smart phones, etc.) are not without design or intention. For instance Apple iOS and Google Android OS both only support a sub-section of Unicode. Therefore any text input solution outside of the supported options on the OS is not an option to endusers. (Not to say that special apps can't be developed, only that OS supported solutions are not available out of the box.) So there are specific limitations based on the device's default operating system.

Target market is a design consideration. This means that the orthographies and text input requirements of target markets hold a higher developmental value in the creating of devices, than design requirements which stand 'out of scope'.

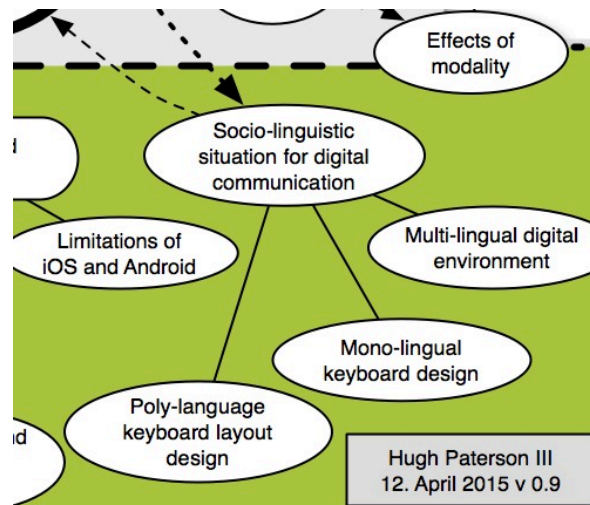
Device based design

A smart phone's build in text input solution is fundamentally different from a standard keyboard. A smart phone with a touch screen has the option present a user auto suggestions. It is also unencumbered by a physical arrangement of keys or characters

on key-tops. This created a level of flexibility and variation in text input solutions. This means that designers can take advantage of options in the hardware which are not available on other hardware options.

Socio-linguistic context

The final section of the diagram is relevant to the language use situation of the device end user. Some text input solutions (including keyboards) are designed for a



single language. The assumption about the user is that they are in a mono-lingual environment or operate mainly in one language, with occasional words or character needs from other languages. However, this is not true in all parts of the world. For instance in Germany, in business situations people may operate in a German primary environment but need to write the occasional email in French or English. Similarly in Francophone Africa users may use any of a variety of languages. They may use a minority language to speak with family, a language of wider communication to speak with business colleagues and they may use French to speak with the government. This context carries over into digital environments as well. In Mexico, the Me'phaa speaker

might have a Spanish working environment or they might have a Me'phaa working environment.

The design criteria is different when the end user is primarily coming from a mono-lingual context as apposed a multi-lingual context. Text input solutions which need a single character of rare use to accommodate a rarely used language would evaluate and place that character differently than when there is a frequent character used in one of two primary languages that the end user is assumed to use.

