Making the 'Next Billion' Demand Access

The Local-Content Effect of `google.co.za` in Setswana

Bastiaan Quast

Abstract

This paper shows that an exogenous increase in accessibility of local language content leads to an increase in demand for internet connectivity among native speakers. Internet connectivity provides enormous improvements in quality of life as well as opportunities for the newly connected, yet recent attempts to connect the current 'next billion' in places such as sub-Saharan Africa have not met expectations. In places where infrastructure has come online and prices have gone down, the expected consequent increase in usage was not observed. The introduction of the Setswana language in the South-African Google Search website was a spillover of the translation from English to Setswana being done for the Botswanan Google Search website. This exogenous improvement of the availability of Setswana-language content has resulted in a substantial increase in the number of native Setswana speakers coming online and owning personal computers. It has also led to increased usage of the Setswana language online. This suggests that connecting the fourth billion will require a greater focus on demand by mean of local content.

1 Introduction

The term 'Connecting the Next Billion' was introduced in The Economist's 2006 'End of Year Report' (Standage, 2006), since then, close to 2 billion people are estimated to have been connected to the internet, up from the just over one billion at the time of writing (Sanou, 2015). However, it is unlikely that the current 'Next Billion' will be connected as easily as the previous ones. In the period 2010-2014 the average year-on-year growth of internet bandwidth in sub-Saharan was around 52%. The increased bandwidth also caused a corresponding drop in the price of internet access. The sub-Saharan average of a 500MB prepaid internet bundle came down to around $10, putting it within range of emerging middle classes in many of these countries. Yet, despite increased range and improved affordability, sub-Saharan Africa is showing stagnation in the growth of internet connections.

As is shown in Figure 1, mobile internet growth, the oft mentioned key driver of internet accessibility in sub-Saharan African, is showing stagnation to the point of being near level or slightly negative. Unlike in other regions in the figure, this observed stagnation is not a consequence of near saturation, as adoption levels in 2015 are only just over 60% on average.

Figure 1: Mobile Internet Growth
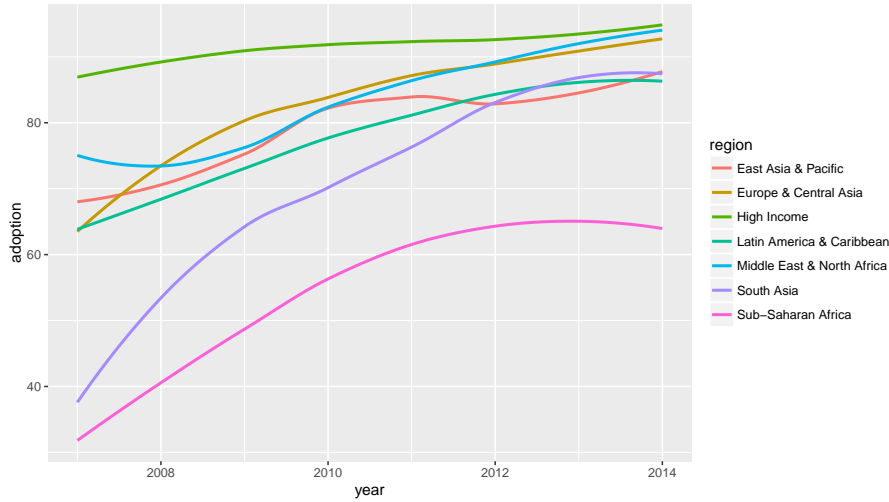
This paper proposes that local content will be vital in bringing the current 'Next Billion' future internet users online. The current offering of online content, primarily in small number of languages, provides insufficient incentive for many to seek access to the internet. In Figure 2, we can see how certain languages, primarily English, represent a relatively small number of native speakers ~10%, but already a much larger number of internet-connected global population, ~25%, but represent more than 50% of all the online content. The 'other'

languages in this figure, represent over half the global population, as can be seen in the top bar, yet all of these languages combined only account for around 10% of online content.

Figure 2: Native Speakers and Web Content
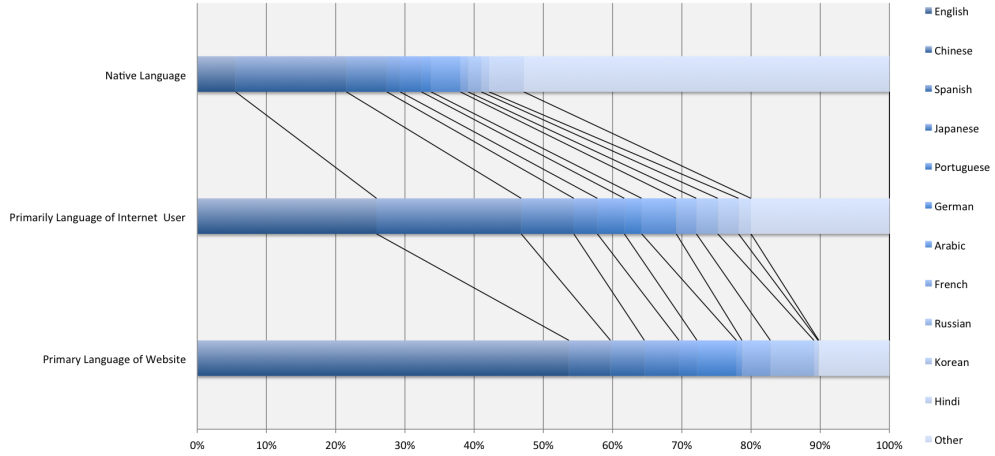
This paper exploits and exogenous increase in the availability of local content in the Setswana language in South African, in order to isolate the increase of internet usage among native speakers. In 2010 Google collaborated with a Botswanan team of linguists (Otlogetswe, 2010) to make its Botswanan website (`google.co.bw`) available in the local language: 'Setswana'. In addition to being spoken in Botswana, there is also a sizable population of Setswana directly across the border in South Africa, where it is also one of the official state languages. This therefore also led to the introduction of the Setswana language on the South African Google website (`google.co.za`), as spillover of the translation work performed for Google's Botswanan website. This change led to a large increase in the number of native Setswana speakers reporting to have spent some amount of money in the past 30 days on internet access.

The Google Search interface represent a very small number of words on the internet and it is not required to use a certain interface language in order to search for content in this language. However, the search page is in many cases the first website viewed by users. Asides from being able to understand the interface of the website, having this interface be in a local language also encourages usage of this local language, which in turn reveals more local language content.

In short, we can identify several major channels which promote increase online engagement, which together constitute the theory of change. Firstly, the ability to read and understand the words of the interface increases the chance that user continues using the website and the internet at large. Secondly, the visibility of local language content increases the likelihood of the user entering

3

search queries in the local language and thereby finding more content in the local language.

The vast majority of internet access is developing countries is through handheld devices such as smartphones. However, due to the limited 'real estate' (surface area of the screen) on a mobile website, the link to changing the interface language is replace with a dropdown menu that reveals the additional language options. Generally, the website will default the the operating system (Android / iOS) language, however, since many local African languages are not available as a system language, this is not possible there. The fact that the introduction seems to benefit desktop usage but not mobile usage is further substantiated by our results that isolate an increase in computer ownership, but no increase in cell phone ownership.

In addition to the increase in the number of individuals spending on an internet connection, we also find a positive effect on the number of individuals living in households with a computer.

The data used for this study comes from the South African National Income Dynamics Survey, provided by Southern Africa Labour and Development Research Unit (2008, 2012, 2013), the data is further discussed in section 2. After whichsection 3 discusses the methods employed in this study, specifically, the discussion of the identification strategy can be found is subsection 3.1 and the use of the Difference-in-Differences estimator in subsection 3.2.

2   Data

South Africa's National Income Dynamics Survey collect data on a representative set of around 10,000 households over time. The first survey took place in 2008, the second one in late 2010 and early 2011, and third one took place in 2012 (Southern Africa Labour and Development Research Unit, 2008, 2012, 2013). It contains an extensive household questionnaire, which breaks household expenditure down into many forms of food and non-food expenditure. In addition to this, the household income is calculated and imputed with other income such as home ownership. The individual (adult) questionnaires also contain information on linguistic skill in English and in the native language, as well as a series of variables relating to communication technology ownership and utilisation. In Figure 3 we can see the number of native speakers of each in the dataset, coloured by the sex of the individual. The dataset contains 51,612 observations (adult individuals), of which 2,806 are female native Setswana speakers and 2,140 are male native Setswana speakers.

A detailed breakdown of the mean number of households that spend on internet access and cellphone usages, as well as cellphone and computer ownership by language and time period (wave) is presented in Table 5 of Appendix B.

4

Figure 3: Native Language and Sex

```
ggplot(adulthh, aes(x = a_lng, fill = a_woman)) + geom_bar()
```
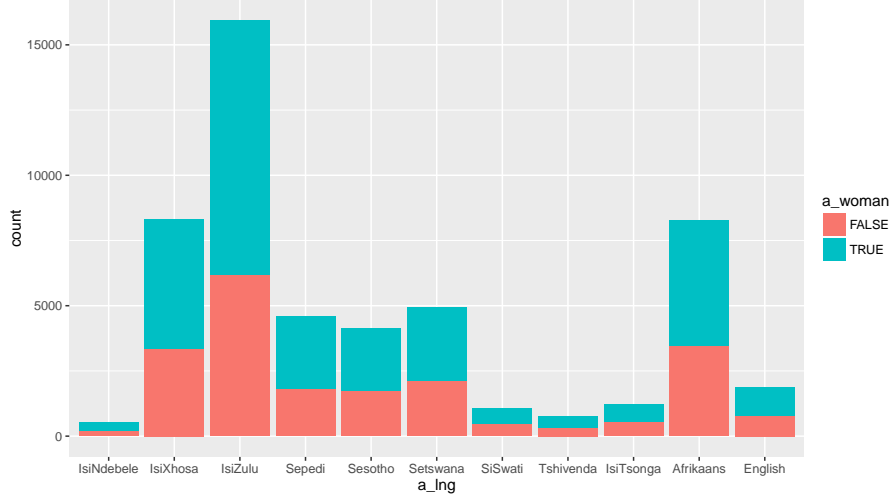

3 Methods

This section beings with a discussion of the identification strategy employed, followed by a explanation of the estimator used to operationalise this, and concludes with a description of the software used for this estimation.

3.1 Identification Strategy

This paper exploits the introduction of the Setswana interface language to Google Search in South Africa as a spillover of the development of that interface for the Botswanan Google Search website. By comparing the number of native Setswana speakers in South Africa being internet users, with the number of South Africans with a different native language around the same time, we isolate the effect of this introduction.

The Setswana language was first developed for the Botswanan Google Search website (`google.co.bw`). As such, the introduction of Setswana to the South African Google Search (`google.co.za`) was a spillover effect of that development. This allow us to rule out any possible endogeneity issues that might otherwise arise in context such as these. For instance, the Afrikaans language is almost solely spoken in South Africa. When we observe that the introduction of the Afrikaans Google Search interface occurs around the same time as a growth in the number of native Afrikaans internet users, it will be hard to isolate the

effect from the introduction from its cause (since an increase in native Afrikaans internet users would be a good reason to introduce it as an interface language).

Substantial numbers of Setswana speakers exist in Botswana, South Africa, Zimbabwe, and to some extend Namibia. However the language is most important in Botswana, where it is spoken by approximately 80% of all people, and where it is the only official language other than English. As such, it is also the place where most linguistic work on the Setswana language takes place. The Setswana Google Search interface was also developed at the University of Botswana by prof. Otlogetswe.

It is worth noting that it is very common not to personally own a computer and 'paying for internet access' therefore also includes a lot of people who use the internet in other locations such as internet cafe's.

In addition to using the propensity to spend on internet (in the last thirty days), we also use the propensity to own a computer as a dependent variable.

3.2 Estimation

As mentioned in the above section, we compare the change in the level of internet users among native Setswana speakers in South Africa, with that of native speakers of other language in South Africa around the introduction of the Setswana interface to the South-African Google Search. For this we use a Difference-in-Differences estimator using a native-Setswana speaker dummy variable (`setswana_logical`), interacted with a event dummy variable (`interface_intro`). The former is `TRUE` when the native language of the individual (`a_lng`) is `Setswana` and `FALSE` otherwise. The later is `FALSE` for data collected prior to the introduction of the Setswana interface language (late 2010, here wave 1 and 2) and `TRUE` after this introduction (here wave 3).

$$y_{irt} = \gamma_{ir} + \lambda_{it} + \delta D_{rt} + \zeta X_{irt} + \epsilon_{irt} \tag{1}$$

In addition to this estimation, we use an alternative specification whereby a factor variable of native language is interacted with the event dummy variable. In a linear regression context, factor variables are estimated as dummy variables for all levels (here: all languages) except for once 'base' level, which is where all language dummies are `FALSE` (i.e. `0`) and the level (native language) thus has to be the first one (here `IsiNdebele`).

3.3 Software

In order to make the result as easily reproducible as possible, this research and writing in the article has been done exclusively using open-source software such as R (R Core Team, 2016). This document is written and LyX (LyX Team, 2016) in the LaTeX(Lamport, 1985) language and compiled using the LuaTeX implementation(Hoekwater et al., 2016). The integration of R code in the document is performed using the knitr implementation (Xie, 2015) of the Sweave framework (Leisch, 2002).

All changes are logged using the version control system Git (Git Team, 2016) and publicly available on GitHub at https://github.com/bquast/Making-Next-Billion-Demand-Access/[1]

4 Results

In the base model, we use an interaction of the `interface_intro` dummy and `setswana_logical` dummy in order to isolate the effect on the explanandum, a dummy variable describing household expenditure on internet in the last thirty days or not (`h_nfnet`, household non-food internet). The results of this estimation are presented in Table 1.

We find that the the interaction term of the event dummy (`interface_intro`) and the native Setswana speaker dummy (`setswana_logical`) is positive and highly significant, with a p-value around `0.0018`. Both the individual dummy variables (`interface_introTRUE` and setswana_logicalTRUE) yield significant but negative parameter estimates. In addition to this, the covariates included in the estimation are also highly significant. The highest education level of the individual (`best_edu`) and the household income (`hhincome`) are both positive and significant. The parameter estimate of `a_womanTRUE` here is negative but not at all significant, this is unsurprising as we use internet expenditure at a household level. Most women live in a household which includes men and visa versa, suggesting that this effect cannot be isolated in this estimation. We further investigate this issue in a separate estimation discussed below. The variables describing linguistic skills in reading and writing in both English and the native language do yield many significant results, though lower levels of English writing skill seems to be correlated with a lower propensity to use the internet (`a_edlitwrten` for levels `2` and `3`, but not the very lowest: `4`).

In an alternative formulation, we include the native language variable as a categorical variable (`a_lng`), interacted with the `interface_intro` dummy. In this estimation we only find significantly positive results for `Setswana` and `Venda` (as small language from the north-eastern region), and a significantly negative effect for the language `Afrikaans`.

---

[1]The repository can by cloned to a local computer by entering in following command in a terminal with Git:
`git clone https://github.com/bquast/Making-Next-Billion-Demand-Access.git`

Table 1: Propensity to live in Household that Spent on Internet in Last 30 Days

```r
lm(h_nfnet ~   interface_intro*setswana_logical +
               factor(a_edlitrden)                    +
               factor(a_edlitwrten)                    +
               factor(a_edlitrdhm)                     +
               factor(a_edlitwrthm)                    +
               a_woman                                 +
               hhincome                                +
               best_edu)
```

	Estimate	Std. Error	t value	Pr($>$|t|)
(Intercept)	-0.0008565	0.0018351	-0.4667340	0.6406924
interface_introTRUE	-0.0117376	0.0012170	-9.6449716	0.0000000
setswana_logicalTRUE	-0.0135286	0.0024255	-5.5777454	0.0000000
factor(a_edlitrden)2	0.0015879	0.0040272	0.3942904	0.6933684
factor(a_edlitrden)3	0.0002543	0.0052755	0.0481963	0.9615600
factor(a_edlitrden)4	-0.0036574	0.0068160	-0.5365861	0.5915561
factor(a_edlitwrten)2	-0.0101933	0.0040178	-2.5369998	0.0111839
factor(a_edlitwrten)3	-0.0107676	0.0052214	-2.0621916	0.0391951
factor(a_edlitwrten)4	-0.0071685	0.0066937	-1.0709421	0.2842010
factor(a_edlitrdhm)2	-0.0025418	0.0036848	-0.6898084	0.4903181
factor(a_edlitrdhm)3	-0.0028899	0.0051291	-0.5634293	0.5731453
factor(a_edlitrdhm)4	-0.0079926	0.0069070	-1.1571655	0.2472107
factor(a_edlitwrthm)2	0.0008117	0.0037229	0.2180373	0.8274010
factor(a_edlitwrthm)3	0.0013993	0.0051372	0.2723915	0.7853222
factor(a_edlitwrthm)4	-0.0069174	0.0069137	-1.0005354	0.3170567
a_womanTRUE	-0.0013881	0.0011452	-1.2121278	0.2254696
hhincome	0.0000027	0.0000001	46.2987087	0.0000000
best_edu	0.0013582	0.0001164	11.6710813	0.0000000
interface_introTRUE:setswana_logicalTRUE	0.0119717	0.0038666	3.0961443	0.0019617

Furthermore, we also use a variant of the base model, in which the propensity of adults (`a_owncom`) to own a computer is used as an explanandum. This is of particular relevance, as the explanandum here (`a_owncom`) differs from the base model's explanandum in two ways. Firstly, it does not include expenditure on internet in ways such as internet cafes, focusses on actual ownership, signaling a more long-term investment and interest. Secondly, the `h_nfnet` variable is at a household level, whereas the `a_owncom` variable is at the level of an individual adult. The results from this estimation are included in Table 2. This form of the estimation yields similar results to those estimated in the base model. Firstly we find that the variable of interest, the interaction term between the event and the language dummy (`interface_introTRUE:setswana_logicalTRUE`) is positive and highly significant, with a p-value smaller than `0.001`. The indi-

vidual dummy variables (`interface_introTRUE` and `setswana_logicalTRUE`) again are significant and negative with for the former's p-value smaller than `0.01` and the later's small than `0.001`. In terms of the linguistic skill, we find that the lower levels of English reading as well as English writing are correlated with lower propensities of computer ownership. Similar to internet expenditure model, household income (`hhincome`) and highest level of education (`best_edu`) are both positive and highly significant (p-value: `~0`). However unlike in the household internet expenditure model, the sex of the individual here is highly significant, specifically, parameter estimate of a_womanTRUE is negative and highly significant (p-value: `~0`). As mentioned above, this variable is difficult to interpret when using a household-level variable as an explanandum, however, here, the computer ownership variable is at an individual level, makes the coefficient more interpretable.

Table 2: Computer in Household

```
lm(a_owncom ~ interface_intro*setswana_logical +
          factor(a_edlitrden)                    +
          factor(a_edlitwrten)                   +
          factor(a_edlitrdhm)                    +
          factor(a_edlitwrthm)                   +
          a_woman                                +
          hhincome                               +
          best_edu)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0075393	0.0031352	2.4047103	0.0161891
interface_introTRUE	-0.0053820	0.0020917	-2.5729824	0.0100856
setswana_logicalTRUE	-0.0147502	0.0041653	-3.5411916	0.0003987
factor(a_edlitrden)2	-0.0306749	0.0069077	-4.4406931	0.0000090
factor(a_edlitrden)3	-0.0310090	0.0090330	-3.4328617	0.0005978
factor(a_edlitrden)4	-0.0389174	0.0116674	-3.3355679	0.0008519
factor(a_edlitwrten)2	-0.0173238	0.0068925	-2.5134122	0.0119602
factor(a_edlitwrten)3	-0.0210640	0.0089384	-2.3565895	0.0184476
factor(a_edlitwrten)4	-0.0187291	0.0114540	-1.6351572	0.1020227
factor(a_edlitrdhm)2	-0.0017925	0.0063219	-0.2835349	0.7767681
factor(a_edlitrdhm)3	-0.0041655	0.0088001	-0.4733526	0.6359638
factor(a_edlitrdhm)4	-0.0267964	0.0118797	-2.2556443	0.0240974
factor(a_edlitwrthm)2	0.0010896	0.0063817	0.1707311	0.8644360
factor(a_edlitwrthm)3	-0.0020020	0.0088176	-0.2270503	0.8203856
factor(a_edlitwrthm)4	-0.0357886	0.0118921	-3.0094517	0.0026186
a_womanTRUE	-0.0230335	0.0019598	-11.7530143	0.0000000
hhincome	0.0000058	0.0000001	56.8794124	0.0000000
best_edu	0.0058419	0.0002002	29.1761771	0.0000000
interface_introTRUE:setswana_logicalTRUE	0.0238052	0.0066799	3.5637169	0.0003660

Asides from the effect on the proportion of households with internet expenditure, and the proportion of adults who own a computer, we also estimate any possible effects on the propensity of the household to own spend on a cellphone, and on adults propensity to own a cellphone.

We find no significant effects on the propensity of expenditure on cell phones or the propensity to own one. As discussed in the introduction, we suspect this to be a consequence of the fact that language switching on mobile cannot be automatic, since the Android operating system does not support the Setswana language, combined with the fact that the Setswana interface button is not visible directly on the `google.co.za` homepage, but rather in a dropdown menu (Figure 7).

This last effect is also visible in Figure 4, which illustrates how usage of the

Setswana word 'thuso', meaning 'help' in search queries, was zero prior to the introduction of the Setswana search interface and became widespread thereafter.

Figure 4: Usage of Setswana Words on Google.co.za

```
ggplot(thuso) + geom_line(aes(x = date, y = thuso))
```
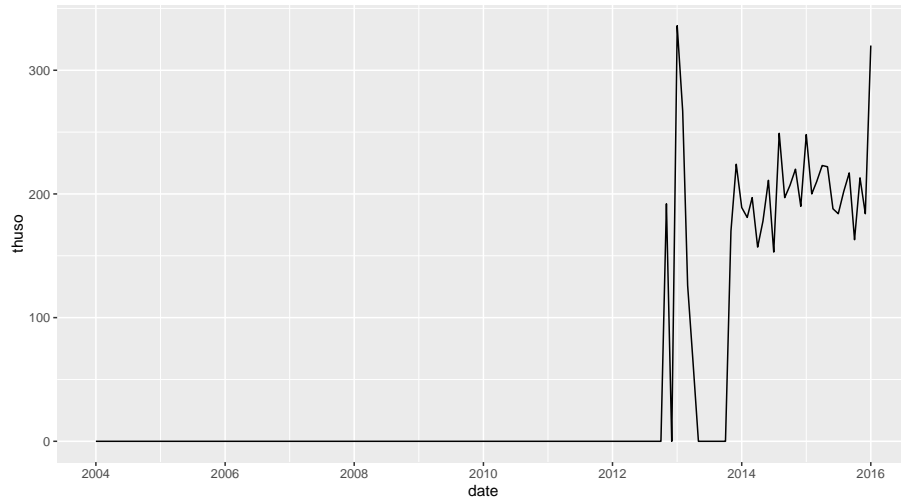

5   Conclusions and Limitations

The vast increase of internet usage among the Setswana speaking population as a result of the newly introduced interface language on `goog.co.za`, suggest that there is a serious lack in the availability of local content in many African languages, which serves as an impediment to further internet adoption here. Additionally, there was also an observably positive effect on computer ownership amongst adults. This suggest that the effect is unlikely to be ephemeral in nature, since computer ownership constitutes a more long-term investment in internet access.

Finally, the fact that this effect is not observed in cellphone ownership and expenditure, which is possibly a consequence of the fact that the interface cannot automatically switch and that the link 'Setswana' is not directly visible from the landing page, suggests that it is important either make the link directly visible on the landing page.

11

References

Git Team

    2016   *Git: Software Code Manager*, 137 Montague ST STE 380, Brooklyn, NY 11201-3548, `http://www.git-scm.org/`.

Hoekwater, Taco, Hartmut Henkel, and Hans Hagen

    2016   *LuaTeX*, `http://www.luatex.org/`.

Lamport, Leslie

    1985   *I1 (\ LaTeX)—A Document*, pub-AW, vol. 410.

Leisch, Friedrich

    2002   "Sweave: Dynamic generation of statistical reports using literate data analysis", in *Compstat*, Springer, pp. 575-580.

LyX Team

    2016   *LyX*, Free Software Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301, USA, `http://www.lyx.org/`.

Otlogetswe, Thapelo J.

    2010   "Setswana Google is here!", *T.J. Otlogetswe Blog*, `http://otlogetswe.com/2010/08/13/setswana-google-here/`.

R Core Team

    2016   *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, `http://www.R-project.org/`.

Sanou, Brahima

    2015   "The World in 2015: ICT facts and figures", *International Telecommunications Union*.

Southern Africa Labour and Development Research Unit

    2008   *National Income Dynamics Study, Wave 1*, version 5.3, `http://www.nids.uct.ac.za/home/`.

    2012   *National Income Dynamics Study, Wave 2*, version 2.3, `http://www.nids.uct.ac.za/home/`.

    2013   *National Income Dynamics Study, Wave 3*, version 1.3, `http://www.nids.uct.ac.za/home/`.

Standage, Tom

    2006   "Connecting the next billion", *The Economist-The World in 2006*, p. 117, `http://www.economist.com/node/5134746`.

Xie, Yihui

    2015   *Dynamic Documents with R and knitr*, Chapman and Hall/CRC, vol. 29, ISBN: 978-1498716963, `http://yihui.name/knitr/`.

A   Factor vs. Dummy

In addition to estimating our model using a dummy variable for native Setswana speakers, we also estimate the model using a factor variable of the categorical variable describing language. In a linear model this is operationalised as a dummy variable for each level except for the base level (here `IsiNdebele`). The results of this estimation are similar to the base model, suggesting that the results are robust to specification idiosyncrasies.

Table 3: Factor of Language

```
lm(h_nfnet ~ interface_intro*a_lng)
```

	Estimate	Std. Error	t value	Pr($>$|t|)
(Intercept)	0.0153374	0.0068968	2.2238509	0.0261630
interface_introTRUE	-0.0153374	0.0115842	-1.3239954	0.1855107
a_lngIsiXhosa	-0.0119771	0.0071155	-1.6832303	0.0923369
a_lngIsiZulu	-0.0101387	0.0070175	-1.4447738	0.1485278
a_lngSepedi	-0.0118839	0.0073155	-1.6244854	0.1042787
a_lngSesotho	-0.0017102	0.0073335	-0.2331986	0.8156081
a_lngSetswana	-0.0100812	0.0072567	-1.3892233	0.1647711
a_lngSiSwati	-0.0106935	0.0084599	-1.2640221	0.2062281
a_lngTshivenda	0.1182366	0.0101757	11.6194711	0.0000000
a_lngIsiTsonga	0.0258487	0.0085505	3.0230556	0.0025036
a_lngAfrikaans	0.0293054	0.0071163	4.1180475	0.0000383
a_lngEnglish	0.0884438	0.0077486	11.4141745	0.0000000
interface_introTRUE:a_lngIsiXhosa	0.0139527	0.0119317	1.1693853	0.2422541
interface_introTRUE:a_lngIsiZulu	0.0146315	0.0117708	1.2430305	0.2138625
interface_introTRUE:a_lngSepedi	0.0169315	0.0122000	1.3878269	0.1651960
interface_introTRUE:a_lngSesotho	0.0130134	0.0122768	1.0600000	0.2891498
interface_introTRUE:a_lngSetswana	0.0207076	0.0121639	1.7023816	0.0886902
interface_introTRUE:a_lngSiSwati	0.0241705	0.0141419	1.7091434	0.0874307
interface_introTRUE:a_lngTshivenda	-0.1182366	0.0154930	-7.6316105	0.0000000
interface_introTRUE:a_lngIsiTsonga	-0.0234958	0.0140080	-1.6773078	0.0934887
interface_introTRUE:a_lngAfrikaans	0.0106404	0.0119382	0.8912936	0.3727760
interface_introTRUE:a_lngEnglish	0.0139328	0.0132294	1.0531742	0.2922663

Lastly, we also estimate the factor model with the inclusion of the linguistic skill variables. These results are again similar to the ones from using simply a dummy variable for native Setswana speakers, suggesting robustness to specification idiosyncrasies.

Table 4: LM4_1: with read / write in eng / native and woman

```
lm(h_nfnet ~   interface_intro*a_lng    +
               a_edlitrden              +
               a_edlitwrten             +
               a_edlitrdhm              +
               a_edlitwrthm             +
               a_woman)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0317749	0.0071775	4.4270461	0.0000096
interface_introTRUE	-0.0160663	0.0115540	-1.3905423	0.1643709
a_lngIsiXhosa	-0.0099180	0.0072875	-1.3609631	0.1735320
a_lngIsiZulu	-0.0074551	0.0071962	-1.0359807	0.3002165
a_lngSepedi	-0.0117930	0.0074983	-1.5727489	0.1157838
a_lngSesotho	-0.0024452	0.0074934	-0.3263203	0.7441835
a_lngSetswana	-0.0090956	0.0074396	-1.2225899	0.2214908
a_lngSiSwati	-0.0109608	0.0086683	-1.2644745	0.2060661
a_lngTshivenda	0.1112485	0.0102479	10.8557149	0.0000000
a_lngIsiTsonga	0.0241788	0.0088017	2.7470744	0.0060153
a_lngAfrikaans	0.0327886	0.0072883	4.4988214	0.0000068
a_lngEnglish	0.0833091	0.0078893	10.5597493	0.0000000
a_edlitrden	-0.0034523	0.0022378	-1.5427211	0.1229053
a_edlitwrten	-0.0067086	0.0021968	-3.0537886	0.0022610
a_edlitrdhm	0.0005400	0.0021481	0.2513920	0.8015121
a_edlitwrthm	0.0021310	0.0021519	0.9903130	0.3220263
a_womanTRUE	-0.0012363	0.0011577	-1.0678587	0.2855898
interface_introTRUE:a_lngIsiXhosa	0.0132544	0.0118983	1.1139774	0.2652947
interface_introTRUE:a_lngIsiZulu	0.0145905	0.0117427	1.2425235	0.2140498
interface_introTRUE:a_lngSepedi	0.0177767	0.0121732	1.4603215	0.1442085
interface_introTRUE:a_lngSesotho	0.0129867	0.0122347	1.0614651	0.2884841
interface_introTRUE:a_lngSetswana	0.0200827	0.0121370	1.6546587	0.0980005
interface_introTRUE:a_lngSiSwati	0.0253579	0.0141092	1.7972623	0.0723005
interface_introTRUE:a_lngTshivenda	-0.1117493	0.0153827	-7.2646212	0.0000000
interface_introTRUE:a_lngIsiTsonga	-0.0181667	0.0140090	-1.2967938	0.1947086
interface_introTRUE:a_lngAfrikaans	0.0095960	0.0119027	0.8062092	0.4201264

B   Ownership and Expenditure by Native Language

The below table breaks down computer and cellphone ownership as well as internet and cellphone expenditure by linguistic group.

Table 5: Descriptive statistics on Ownership and Expenditure

```
adulthh %>%
  group_by(a_lng, wave) %>%
  summarise(a_owncel   = mean(a_owncel,   na.rm = TRUE),
            a_owncom   = mean(a_owncom,   na.rm = TRUE),
            h_nfcel    = mean(h_nfcel,    na.rm = TRUE),
            h_nfnet    = mean(h_nfnet,    na.rm = TRUE),
            h_nfcelspn = mean(h_nfcelspn, na.rm = TRUE),
            h_nfnetspn = mean(h_nfnetspn, na.rm = TRUE))
```

a_lng	wave	a_owncel	a_owncom	h_nfcel	h_nfnet	h_nfcelspn	h_nfnetspn
IsiNdebele	1	0.6026490	0.0331126	0.6533333	0.0000000	59.78519	0.0000000
IsiNdebele	2	0.6864865	0.0270270	0.6800000	0.0284091	84.12571	1.4204545
IsiNdebele	3	0.7611111	0.0333333	0.8722222	0.0000000	105.46067	0.0000000
IsiXhosa	1	0.4631115	0.0112269	0.4352518	0.0012043	39.76659	0.0574297
IsiXhosa	2	0.6065066	0.0186335	0.5996664	0.0054517	54.47248	0.2978972
IsiXhosa	3	0.7564988	0.0345508	0.7041694	0.0019756	86.33322	0.0869279
IsiZulu	1	0.5610984	0.0132693	0.5756053	0.0013730	48.21557	0.0363844
IsiZulu	2	0.5169004	0.0127744	0.4721318	0.0086366	49.00402	1.3053671
IsiZulu	3	0.7662405	0.0252420	0.7247881	0.0044928	96.59072	0.6889696
Sepedi	1	0.6127214	0.0265273	0.5772947	0.0048309	40.47351	0.3462158
Sepedi	2	0.7029372	0.0226818	0.5933485	0.0021994	51.69161	0.1583578
Sepedi	3	0.7936063	0.0718697	0.7863152	0.0050477	110.64450	1.1385306
Sesotho	1	0.6562986	0.0366044	0.5507812	0.0062598	49.45647	0.4772727
Sesotho	2	0.6973684	0.0457010	0.5411671	0.0213640	60.29976	3.0657354
Sesotho	3	0.8114210	0.0949535	0.8147901	0.0113032	115.52993	1.8284574
Setswana	1	0.5796003	0.0351724	0.5281593	0.0068681	59.33404	0.2886598
Setswana	2	0.6004872	0.0359537	0.6494778	0.0037783	86.17885	1.0214106
Setswana	3	0.7728036	0.0711086	0.7684564	0.0106264	112.96165	0.8509804
SiSwati	1	0.6593060	0.0441640	0.7823344	0.0000000	59.17647	0.0000000
SiSwati	2	0.7598870	0.0612813	0.6693227	0.0091185	63.28685	0.4589666
SiSwati	3	0.8247978	0.0458221	0.7816712	0.0134771	120.19944	1.4555256
Tshivenda	1	0.5980861	0.0334928	0.5645933	0.0000000	108.45631	0.0000000
Tshivenda	2	0.7804878	0.0000000	0.9621212	0.5441176	50.85606	1.2058824
Tshivenda	3	0.8617363	0.0225080	0.8456592	0.0000000	142.95035	0.0000000
IsiTsonga	1	0.6411765	0.0235294	0.4408284	0.0000000	27.80896	0.0000000
IsiTsonga	2	0.7375887	0.0118203	0.7163636	0.0929368	55.88364	0.9368030
IsiTsonga	3	0.8621495	0.0397196	0.8691589	0.0023529	103.93128	0.3529412
Afrikaans	1	0.5392884	0.1345441	0.6227876	0.0465116	133.43521	9.4163569
Afrikaans	2	0.5422477	0.0904605	0.6258591	0.0424710	103.81572	10.9746890
Afrikaans	3	0.6686971	0.1106225	0.7645862	0.0399458	146.19560	10.5008501
English	1	0.7266667	0.2969374	0.7449933	0.1016043	371.01291	31.0356653
English	2	0.7976190	0.3234127	0.8728814	0.1070707	375.17797	35.3555556
English	3	0.8608059	0.3156934	0.8811700	0.1023766	377.87127	23.3816514

C   Descriptive Statistics

The Figure 5 describes the skill of individuals in reading the English language, where 1 the best and 4 is the worst, grey values are `NA`.
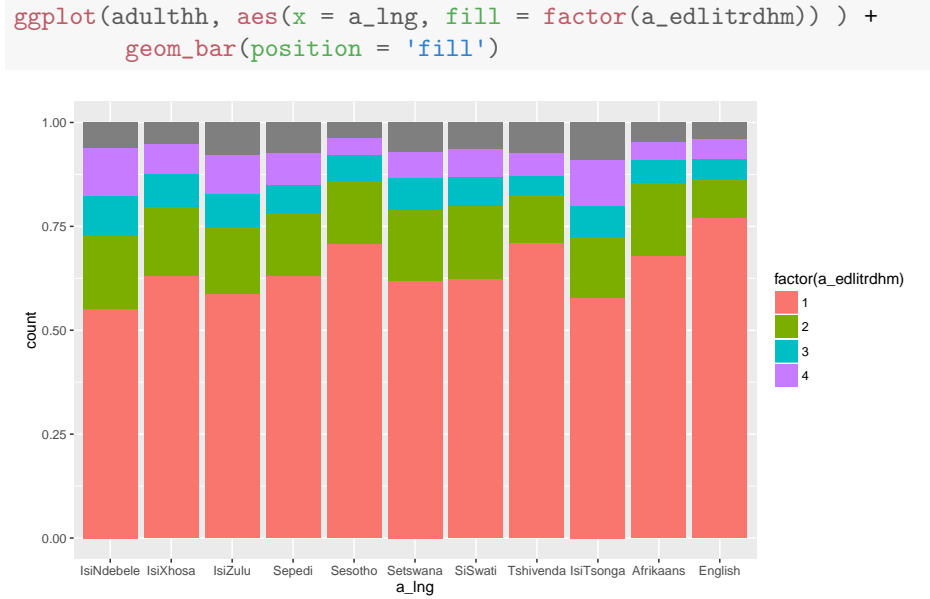
Figure 5: English Language Reading Skills

```
ggplot(adulthh, aes(x = a_lng, fill = factor(a_edlitrden)) ) +
        geom_bar(position = 'fill')
```
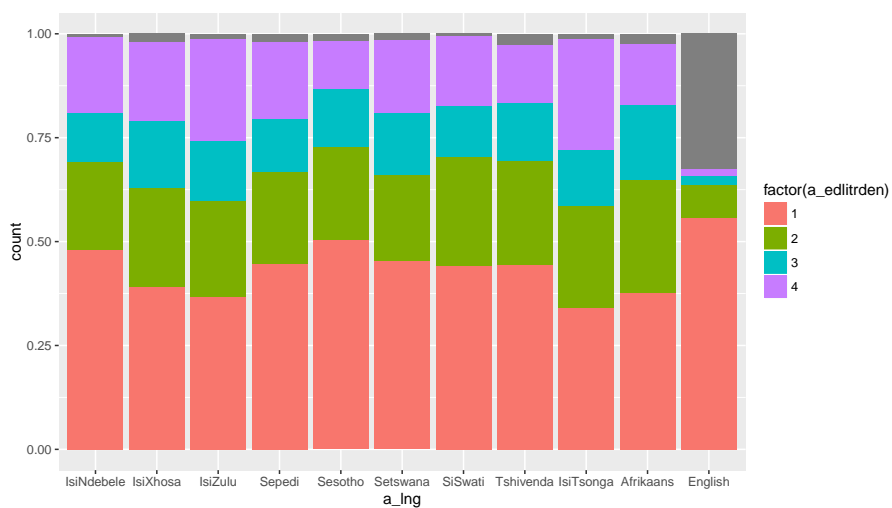

The Figure 6 does the same but with regards to the native language.

Figure 6: Native Language Reading Skills

```
ggplot(adulthh, aes(x = a_lng, fill = factor(a_edlitrdhm)) ) +
        geom_bar(position = 'fill')
```


D   Language Switching on Mobile

The graphics in Figure 7 illustrate the process of switching the mobile `google.co.za` interface language to Setswana, from the default English. The mobile interface language can automatically be changed based on the system language of the operation system (in most cases Android), however, as Setswana and many African languages are not available as system languages in Android, the website interface on `google.co.za` will default to English.

Figure 7: Changing Interface Language on Mobile