UJ Sociology, Anthropology & Development Studies

W E D N E S D A Y S E M I N A R

Hosted by the Department of Sociology and the
Department of Anthropology & Development Studies

UNIVERSITY OF JOHANNESBURG

Meeting no 3/2010

To be held at 15h30 on Wednesday, 10 February 2010,

in the Anthropology & Development Studies Seminar Room, DRing 506, Kingsway campus

Languages on the internet:

Reflections on the changing political economy of new media

- Please do not copy or cite without authors' permission -

Dr Lloyd Hill

University of Stellenbosch

- Programme and other information online at www.uj.ac.za/sociology -

Languages on the Internet – reflections on the changing political economy of new media

Lloyd Hill[1]
Department of Sociology and Social Anthropology
University of Stellenbosch

"When I took office, only high energy physicists had ever heard of what is called the World Wide Web... Now even my cat has its own page."

Bill Clinton[2]

Introduction

This is an exploratory paper, intended to serve as a conceptual prelude to a more systematic empirical study of South African 'language domains' on the Internet. The idea for this project has been gestating for some time and, following my recent move to Stellenbosch, will probably require some more time before it is manifested in a substantive study of South African languages. For the purpose of this paper it will therefore suffice to develop a kind of 'problem statement' for the proposed broader study. This will proceed in two stages: (1) an attempt to conceptualize 'languages on the Internet' as a subset of a broader historical-constructivist concept of 'languages'; and (2) a brief discussion of recent trends associated with 'new media' as a precursor to the exploration of two contrasting approaches to the study of 'languages on the Internet.' The first section presents my theoretical treatment of 'language', which is based on elaboration of the concept of 'new media.' The second section then explores two contrasting ways in which 'Internet languages' can be conceptualized and researched. I also discuss the potential to integrate these approaches in a study of South African languages on the Internet. The final section provides a few 'broad brush' comments on the changing political economy of the Internet, in an effort to fuel discussion and generate some hypotheses on the relationship

---

[1] I can be contacted at lloydhill@sun.ac.za. This paper was presented at UJ Sociology, Anthropology & Development Studies Wednesday Seminar on 10 February 2010.

[2] http://www.quotationspage.com/quote/31751.html (accessed 7 February 2010).

between broader Internet developments and language trends in South Africa.

Languages and 'new media'

The argument that follows is premised on a basic distinction between 'language' and 'a language.' The English term (like its equivalent in Afrikaans and many other languages) is ambiguous, representing two concepts that in French are rendered as '*langage*' and '*langue*.' The former refers to language as a universal species capacity (for speech), while the latter refers to specific languages such as English, Afrikaans and Zulu. The tendency to reduce the latter to explanations in terms of the former has a long history in Western scholarship. This tendency is known as Aristotelian surrogationalism and its classical formulation is '*aliquid stat pro aliquo*' or 'spoken words are symbols of mental events, and written of spoken.'[3] My concern is therefore for 'languages' as relatively recent historical constructs or artifacts, which are *real* but not *natural*, to the extent that they cannot be reduced to our common biological capacity to speak. One means of avoiding the surrogationalist trap is to define 'a language' or 'languages' explicitly in terms of media and, moreover, in terms of relatively 'new media.'

In recent decades the term 'new media' has come to refer to a range of integrated electronic media that that have emerged in the wake of the post-war computer revolution. While my intension is to focus on these new forms – and the Internet in particular – my use of the term 'new media' is at this juncture more general: what we call 'modern languages' are largely the product of a series of 'new media' that date back to the institution of modern printing (c.1464).[4] Printing revolutionized the political economy of language use, first in Europe and then throughout the world. From its inception the printing press generated two very distinct markets for the printed word: a 'vernacular' market for pamphlets and brief religious tracts and 'a scholarly' Latin market for books on classical Greek and Roman culture. In Europe the Latin market eventually gave way to new markets for 'national languages.' Printing was therefore instrumental in the creation of a powerful new form of society – the 'nation-state.'

The ostensibly simple process of making exact copies of written material publically available therefore produced profound social changes, which culminated in the creation of mass reading populations or 'publics.' Printing had the effect of radically institutionalizing markets for writing and other products of human cognition. Beginning in the nineteenth century, the emergence of mass reading publics stimulated demand for a new generation

[3] Florian Coulmas, *Language and Economy*, Oxford: Blackwell, 1992, p.107.
[4] Lucien Febvre and Henri-Jean Martin, *The Coming of the Book - The Impact of Printing*, 1450-1800. London: Verso, 1984, p.54.

of 'written' technologies and 'reading' practices. [5] For more than a hundred years new developments in communication technology have occurred in Anglophone countries and these technologies have played a particularly significant role in the ascendance of the United States of America. This process began in the nineteenth century, when the telegraph and the newspaper played a crucial role in integrating disparate territories into the United States.[6] In the twentieth century the status of English was greatly enhanced as a result of the development of the telephone, radio and television. Furthermore, Anglo-American collaboration during the two world wars laid the foundations for the commercial exploitation of a new technology that would subsequently revolutionize the field of telecommunications – the computer.

The computer had a profound effect on the post-war establishment of English as the first truly global language of printing and publishing[7]. Firstly, the computer represents a powerful objectification of rationalism: an effective reification of the mind/body dichotomy in the form of distinct mechanical (hardware) and symbolic (software) systems. While the 'first language' of the computer is the binary system known as machine code, the development of compilable 'higher level' languages soon meant that only a handful of computer programmers required a knowledge of machine code. Upper level programming languages are almost universally English-based, and they have spawned a massive new publication industry in the form of software instruction manuals. Secondly, the advent of the networked computer put Anglophone economies at the forefront of a new communication revolution: the gradual merger of computer-based 'information technologies' (IT) and older communication technologies (notably telephone networks), to produce an entirely new field of information and communication technology (ICT). The result of these new technologies was the creation of new channels of communication (new media) and new forms of communicative space, the most significant of which was the Internet.[8]

---

[5] Here I am using the term in the general sense described above, to include a host of new media and practices. In English these new forms of 'writing' are typically indicated by the use of the suffix '-graph': e.g. telegraphy (writing over distance), phonography (writing with sound); and photography (writing with light).

[6] Benedict Anderson, *Imagined Communities - Reflections on the Origin and Spread of Nationalism,* London: Verso, 1983.

[7] Following Latin and French – see Febvre and Martin, 1984.

[8] Colin Gardner, "English and new media", in *Redesigning English*, Abingdon: Routledge, 2007.

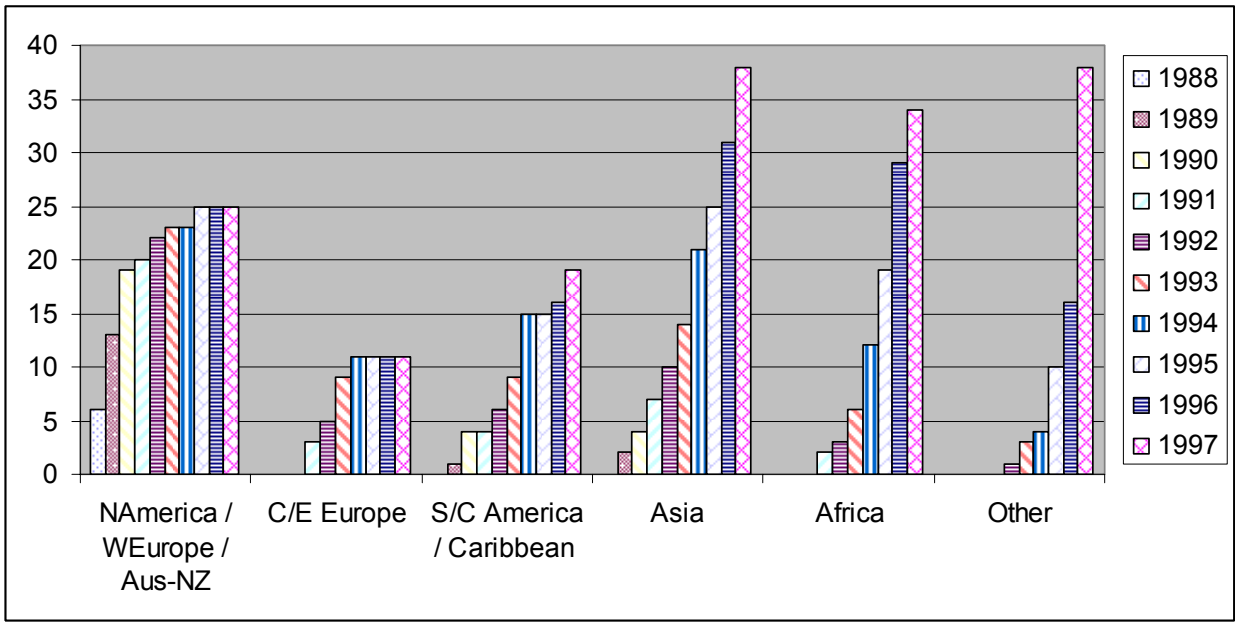
Languages on the Internet

The history of the Internet is commonly traced back to the establishment of the ARPANET in 1969.[9] For reasons related to the Anglo-American dominance in the field of information technology, referred to above, the Internet remained an overwhelmingly Anglophone phenomenon until the 1990s. It is only after the invention of the Hypertext Transfer Protocol (HTTP) in 1990 that the stage was set for the development of a truly global 'World Wide Web.' The invention of the browser in 1993 was the sociological turning point: the graphical interface facilitated the development of a new mass medium, by transcending the language constraints associated with earlier command-line interfaces.

Before the invention of the browser access to the Internet assumed: (a) a minimum basic competence in English; and (b) familiarity with the written and English-based command structures of operating systems and online domains. The development of graphic interfaces for stand-alone computers (e.g. the Windows operating system) and browser-based access to the Web therefore represented a shift towards a hybrid visual-verbal form of online communication, with much greater potential to spread beyond the boundaries of Anglophone economies.

After 1993 we therefore observe two noteworthy social trends: the rapid growth of the number of users situated outside the American 'core' and the rapid connection of countries to the NSFNET infrastructure. The issue of numbers is addressed subsequently. The table below shows the pattern of country connection – a rough indicator of the spread of 'national internets' and more often an indication of registration of a domain name with the Domain Name System (DNS – established in 1984).

[9] In 1958, a year after the USSR launched Sputnik, the United States formed the Advanced Research Projects Agency (ARPA) within the Department of Defence (DoD). The ARPAnet was commissioned by the DoD in 1969. The use of 'Internet' to describe a 'connected set of networks' dates back to the invention of the TCP/IP protocols in 1982 (although the OED gives 1986 as the year in which the capitalized form was first attested). See Hobbes' Internet Timeline 10, http://www.zakon.org/robert/internet/timeline/ (accessed 8 February 2010).

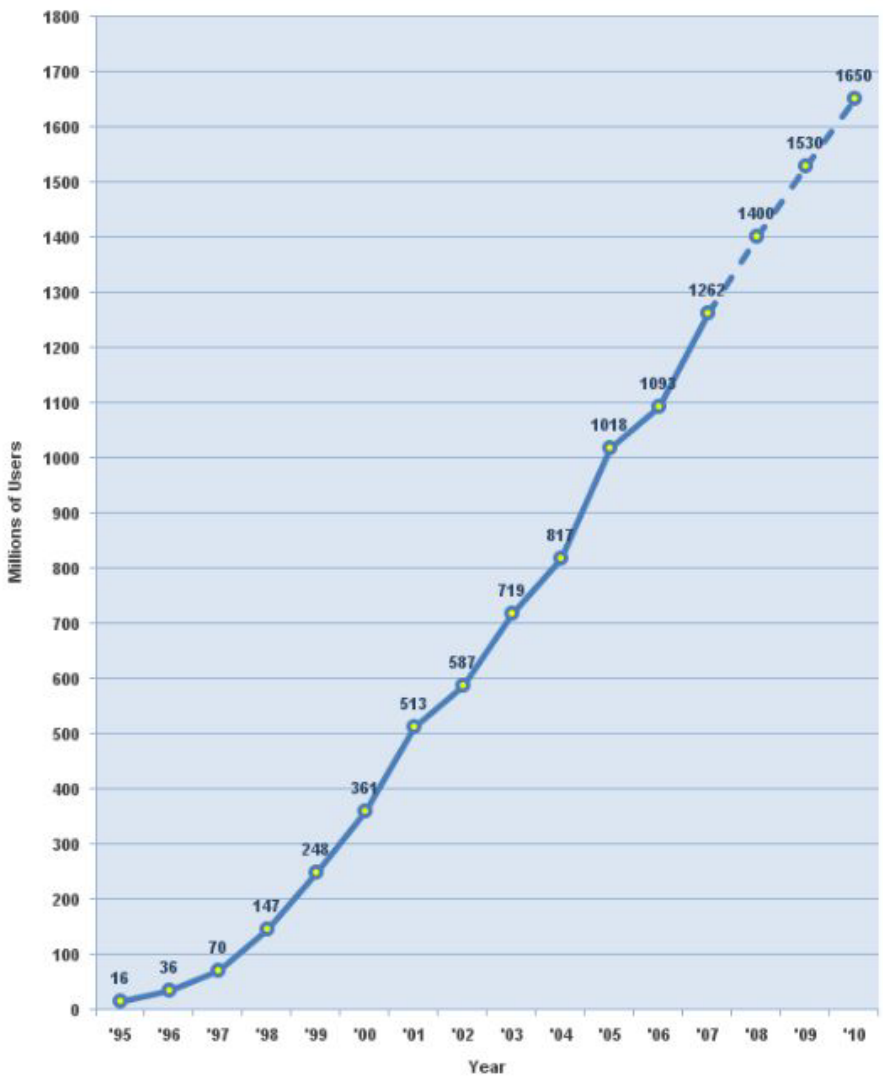


How and to what extent are we able to infer from infrastructure and usage statistics the concurrent trends associated with the spread of 'languages' on the internet? The conceptual model proposed here is built on a broad distinction between 'language speakers' and 'language contexts.' The later can alternatively formulated as the 'contexts of languages' – this broad distinction is premised on the earlier distinction between 'language' and 'a language.' 'Language contexts' can in turn be subdivided into 'technologically mediated situational contexts' and 'symbolic domain.' I argue that it is at the intersection of these two broad sets of contexts that the real challenge lies for researchers interested in 'operationalizing' the study of languages on the Internet.

Two years after the invention of the browser the number of Internet users was estimated to be about 16 million. This figure grew to 36 million in 1996, before increasing tenfold to 361 million by December 2000. The one billion mark was reached towards the end of 2005 and the latest estimate (September 2009) currently stands at 1.734 million.[10]

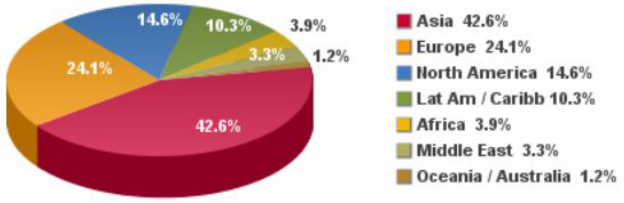
[10] Internet World Stats – usage and population statistics, http://www.internetworld stats.com/emarketing.htm (accessed 8 February 2010).

Internet Users in the World
Growth 1995 - 2010



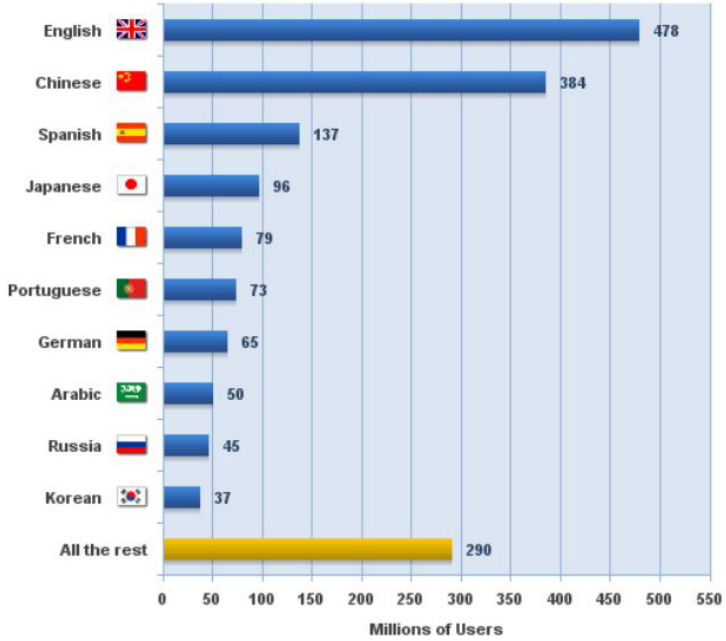
Source: www.internetworldstats.com - January, 2008
Copyright © 2008, Miniwatts Marketing Group

World Internet Users
by World Regions



Source: Internet World Stats - www.internetworldstats.com/stats.htm
1,733,993,741 Internet users for September 30, 2009
Copyright © 2009, Miniwatts Marketing Group

Top 10 Languages in the Internet
millions of users



Source: Internet World Stats - www.internetworldstats.com/stats7.htm
Estimated Internet users are 1,733,993,741 for September 30, 2009
Copyright © 2009, Miniwatts Marketing Group

The statistics provided above suggest that – at least in terms of the people who use it – the Internet is becoming an increasingly multilingual phenomenon. One commentator has noted this change as follows:

> The next billion users will not look like that. They will be a very different group of the people, less willing to speak a common language.[11]

This statement reflects both an important fact and a noteworthy Anglophone bias. It is a fact that the next billion users will demonstrate a very different sociolinguistic profile, but it is misleading to suggest that the issue is willingness to speak a common language. It is far more likely that the next billion users will simply be *unable* to speak English – or rather, unable to use English in the relatively formal and elevated domains of the Internet.

Important as they are as a point of departure, available usage statistics are problematic indicators of online language trends. These problems can be summarized as follows:

- they assume a nominal association between 'a user' and 'a mother tongue', which is frequently tenuous; and
- they tell us something about 'language users', but almost nothing about the patterns of language use and usage.[12]

A conceptually distinct but potentially complementary research objective is therefore the attempt to establish patterns associated with the use of 'languages' in various online contexts. It is furthermore important to recognize that these contexts are not 'flat' as the visual aggregation of user statistics is wont to suggest. Given its historical association with Anglo-American liberalism and a more general 'analytic' orientation to the interpretation of data, there is a popular tendency to present the growth of the Internet as an essentially benign and 'democratic' phenomenon.[13]

The internet is, however, not flat – like other social phenomena it has a hierarchical structure. Recognition of this feature is crucial to studies concerned with the social dimensions or morphology of the Web. The next billion users is emerging from the strata below the global middle class – with no access to the education that guaranteed proficiency in the 'global' medium. Moreover, the reference to 'languages on the Internet' is

---

[11] Ethan Zuckerman, "Over the internet border", *New Scientist,* 2007.

[12] The terms 'use' and 'usage' are commonly used in linguistics to refer to the functional and structural dimensions of language use.

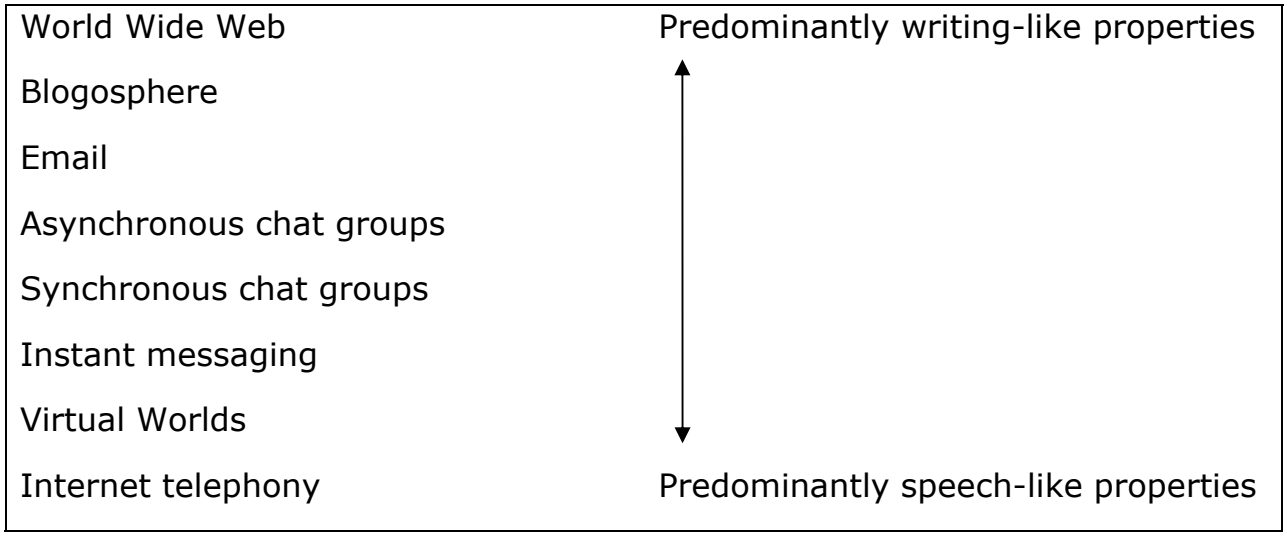
[13] It has, for example, been argued that "the Internet frees us from the limitations of where we're born and where we grew up." See Zuckerman, pp. 42-43.

ambiguous to the extent that it can be measured either in terms of speakers of a given language using the Internet or to language content on the Internet.[14] The table presented above reflects the former and these statistics therefore demonstrate a 'flat' trend: the relative decline of first language English speakers as a proportion of the growing number of global users. This general demographic trend tells us little about the changing pattern of language usage on the Internet, although it does suggest a growing potential demand for multilingual content. An assessment of form and content therefore requires an explicit focus on the intersection between two broad categories of 'context.' Technologically mediated 'contexts of situation'[15] and more abstract 'symbolic domains.'[16]

What type of language medium is the Internet? In a study of language use on the Internet Crystal has identified seven broad 'Internet-using situations': synchronous and asynchronous chat groups; virtual worlds; instant messaging; blogging; electronic mail and the World Wide Web.[17] Since the publication of this book, Internet telephony has come into its own and can therefore be added as an eighth situation. Crystal argues that these contexts of situation can be arranged more or less continuously in terms of the relative speech-like and writing-like properties[18] that they exhibit.

[14] Bram Dov Abrahamson, "Internet Globalization indicators", in *Telecom-munications Policy* 24, 2000, p.70.
[15] 'Context of situation'
[16] 'Symbolic domain'
[17] David Crystal, *Language and the Internet,* Cambridge, UK: Cambridge University Press, 2006, p.11.
[18] Writing tends to be space-bounded, fixed, context independent and highly structured. Speech tends to be time-bound, relatively unstructured and context sensitive. *Ibid.*, p. 28.

Linguistic properties of Internet technologically mediated contexts



Internet telephony is therefore the online situation that most closely resembles everyday speech, while on the other end of the spectrum the World Wide Web and the blogosphere demonstrate many of the properties associated with writing and printing. The fundamental quality of situated speech is its embeddedness in an interactional context and its 'codependence' with other situated semiotic systems: paralinguistic cues, body language, the arrangement of artifacts etc. And these are precisely the qualities that make new Internet contexts accessible to growing numbers of people around the globe. But if speech-like contexts account for the growing popular reach of the Internet, it is writing-like, or rather print-like contexts that account for its tendency towards stratification. Any attempt to include social stratification as a dimension of in the study of online languages must therefore pay careful attention to the structure of online 'symbolic domain.'

'Domain', like 'code', is a term originating in law that has acquired a host of new senses associated with a range of knowledge fields (notably, for our purposes, linguistics[19] and computer science). Moreover, 'domain' has a particular association with geography and the topological arrangement of knowledge more generally. The real challenge in measuring online language trends in a country like South Africa therefore lies in the attempt to explore the relationship between situated access and the more qualitative dimensions of unequal access to content, knowledge environments and web-

[19] In semiotics a domain is a higher level sign, i.e. one that covers or includes other signs. See James P. Spradley (1979) *The Ethnographic Interview*, New York: Holt, Rinehart and Winston, p. 93.

mediated relationships. This sensitivity accounts for the recent emergence of critical orientation towards the concept of a 'digital divide.' Over-optimistic assessments of the egalitarian potential of Internet technologies tend to invest rather heavily in the notion of 'a digital divide' – a concept that has tended to convey a binary distinction between the availability / unavailability of technologically-mediated access[20] to the Internet. What these accounts fail to appreciate is the extent to which the diffusion of the World Wide Web is premised on the cumulative effects of preceding information technologies. Partly through its historical association with telephone networks (and by way of an analogy with the telephone), the predominantly quantitative measures associated with the concept of "digital divide" proved useful during the early stages of Internet diffusion. With the rapid geographical diffusion beyond the boundaries of industrialized states *and the growing penetration* of the Web into the poorer social strata of most societies, it has been argued that the term "digital inequality"[21] is more appropriate, to the extent that it better conveys the technological and social dimensions of Internet access.[22] DiMaggio and Hargittai have explored five aspects of the social dimension of digital inequality:

- Variation in the technical means (hardware / software / connections) by which people access the Internet;
- Variation in the extent to which people exercise autonomy (e.g. working from home) in their use of the Internet;
- Variation in the level of skill that people bring to their use of the Internet;
- Variation in the level of social support available to Internet users; and
- Variation in the purposes for which people use the technology.[23]

A concern for the study of digital inequality is therefore one of the main objectives of the proposed empirical study of South African languages on the Internet; one which follows logically from the aforementioned concern to

[20] Fuchs and Horak note definitions of the digital divide attributed to Manuel Castells ("inequality of access to the Internet") and Jan van Dyk ("the gap between those who do and do not have access to computers and the Internet") before recommending "strategies for closing the global digital divide." Fuchs, C. and E. Horak (2006) "Africa and the digital divide", *Telematics and Informatics* 25, pp. 99-116.

[21] This concept is relatively new and under-theorized. A full-text search for articles in the ScienceDirect online citation index revealed 954 articles using the collocation "digital divide", but only 19 for "digital inequality."

[22] DiMaggio, P. and Hargittai, E. (2001) "From the 'Digital Divide' to 'Digital Inequality': Studying Internet Use as Penetration Increases" Working Paper Series 15, Centre for Arts and Cultural Policy Studies, Princeton University.

[23] DiMaggio and Hargittai (2001).

integrate a study of users with an investigation of the changing structure of online language contexts. It is difficult to state at this juncture how this is to be achieved, but it will no doubt require collaboration between researchers in the fields of sociology/sociolinguistics and information science.

The changing political economy of the Web

I conclude this paper with a number of speculative remarks on the changing political economy of the Internet – my intention being to stimulate debate on the implications that macro-trends hold for future role/status of the Internet in South Africa. The recent dispute between China and Google has put the issue of Internet governance clearly in the public domain. The matter has of course tended to be presented as a 'problem of censorship' , when in fact there are far more fundamental issues at stake. What we witnessing is a stand-off between a waning superpower and a major pretender. With its economic status in rapid decline, the US is increasingly dependent on an awkward amalgam of 'hard' (military) and 'soft' (symbolic) power to sustain its position in the world economy. Central to the latter is its ability to influence (govern?) and increasingly complex and politically fragmented Internet.

Weighing against the continued dominance of the US is the changing structure of Web traffic. Given its origin in the United States, the initial growth of the Internet was centered on this country. Thus studies of geographical bandwidth connectivity have demonstrated the extent to which the internet has been heavily US-centric, "with intraregional traffic in Latin America, Asia, or Europe often going halfway round the world before arriving at its final destination."[24] More recently, the technological "shape" of the Internet has begun to fracture as traffic from other regions has become increasingly less centered on the United States. This is particularly true of European traffic, but East Asia is also emerging as an increasingly distinct hub. In Europe at least two-thirds of international connectivity now remains in the region and European hub cities are becoming increasingly prominent within the global structure.

> Western European cities predominate among the Internet's largest international hub cities. Ten of the twenty largest hub cities are located in West Europe; of them seven are in the top ten, underlining Europe's emergence as a secondary Internet hub – a move away from the US-centric model – but also pointing to the role of European integration, which has translated into lower costs for intra-European connectivity, in fostering this process.[25]

[24] Dov Abrahamson, "Internet Globalization indicators", in *Telecommunications Policy* 24, 2000, p.71
[25] *Ibid.*, p. 73.

Weighing in favour of the US is its continued dominance in the field of information technology and through this its ability to sustain demand for an Anglo-American vision of the future. Google has been very instrumental in this process. On the one hand it has played a pivotal role in the development of 'Web 2' and the individualizing ideal of an "architecture of participation." On the other hand it has been extremely successful in commercializing and stratifying the Web through the development of its search engine and 'Pagerank' technology. Will this continue to unite the Web and if so, how should South Africans position themselves in the changing environment? And how will new developments – notably the so called 'Semantic Web' – change the status of languages on and off the Web? With the exception of Afrikaans, none of the other 'minority languages' in South Africa would seem to have established much of a foothold in writing-like domains of the Internet. To what extent is this a problem, and could changing macro-circumstances provide new opportunities for more and qualitatively different levels of participation?