

**Language, writing and automaticity:
Software studies in front of linguistic capitalism**

Digital Studies Seminar – December 2, 2014

Frédéric Kaplan and Warren Sack

The digital stage of the [grammatization](#) process raises the problem of a radical change in the conditions of reading, writing and linguistic expression. Formalization, discretization and exteriorization of human language behaviors into digital tertiary retentions seem to make impossible the re-appropriation of knowledge by Internet users. Indeed, in order to capitalize on the Internet user's searches, Google exercises a control on language using tools of automatic correction and completion. By encouraging users to use the words which are statistically most frequent and subject of speculation for advertising, these automata bring him into the field of the "predictable" language, commercially exploitable by the company. Thanks to this algorithmic mediation of expression, Google has succeeded in transforming the linguistic material into a genuine economic resource. But this phenomenon, described by Frederic Kaplan under the name "linguistic capitalism", has the direct consequence to regularize and homogenate natural languages and to induce their desidiomatization at a global scale. This retroactive effect of technologies on language seems to lead to the emergence of a new syntax and a new lexicon that are informed by the linguistic capabilities of machines and the economic value of words.

To think and transform this new linguistic environment, it seems necessary to distinguish textual resources that have been produced by humans from those which have been automatically generated by machines, in order to understand the logic at work in the automatic textual production. This is especially what Warren Sack analyzes: if linguistic capitalism currently relies on selling individual words and producing abstracts and translations, it could also be extended to the automatic generation of stories. The development of language planning today allows software to generate automatically stories based on data and programs. Yet, the effort to translate narrative art into computer technology does not simply mechanizes the narrative competence: it implies a re-articulation of narrative logics that modifies the nature of stories and conditions the forms of future literature and fictions.

How to integrate this automatized environment to a project for developing knowledge? This capacity to read and write softwares indeed seems necessary to develop new linguistic capabilities and regain room for maneuver in the exercise of narrative art: not in order to rationalize the existing artistic forms, but to determine the desirable changes and invent new narrative expressions, working against habituation and automation. Software studies could then appear as a mode of resistance to linguistic capitalism: if there is not any fate in the computer implementation of a natural competence but a set of reversible choices, one should define the place where IT and humanities can establish an intelligent dialogue which should benefit all.

F. Kaplan – Language as a capital



[Frederic Kaplan](#) is a Professor at the Ecole Polytechnique Fédérale de Lausanne (chair of digital humanities), and director of the Digital Humanities Lab. He focuses his work on human-machine interfaces and artificial intelligence. In 2011, he publishes in *Le Monde Diplomatique* an [article about the linguistic capitalism](#) Google initiated. He then writes a more academic paper on the subject, published in *Digital Studies. Organologie des savoirs et technologies de la connaissance*. He finally posts on the web an article regarding the linguistic capitalism and the algorithmic mediation. Today, he is working on a study of the internal logic of Google Translate in order to understand the issues of this new economic model and its effects on languages and cultures at a global scale.

After having summarized the five major theses of the linguistic capitalism, Frédéric Kaplan explored the retroactive effects of technologies and economic logics on mother tongues and writing practices. He thus showed that this new linguistic and economic context is at the origin of major linguistic and cultural transformations. Which ones, how?

1/ What is linguistic capitalism?



1/ A new economic regime characterized by the monetization of languages

Linguistic capitalism is a new economic regime initiated by Google and characterized by the commodification of languages at a global scale: the words are given values according through an auction system by advertisers and Google ranking algorithms. There is a words stock exchange and the value of words changes over time (the word “ski” has more value in winter than in summer for example). It is through this commercial model that we have to analyze the innovations and services launched by Google, which gains thanks to this process \$ 50 billion per year (\$ 5 million per hour).

2/ An expression economy and no longer an attention one

Linguistic capitalism is not an [attention economy](#) but an expression economy: the issue is not to attract attention or to influence intentions, but to be a mediator of speech and writing. This model aims at developing intimate and long-lasting linguistic relationships with a great number of users in order to modelize and inflect linguistic behaviors, create a linguistic market and control speculation on words.

3/ The transformation of language into an exploitable economic material

Google follows the movements of living language to optimize it and make it enter into the field of the commercially exploitable static language. Thanks to the tools of [autocompletion](#) for example, a misspelled word without any economic value can be transformed into a potentially profit-making resource.

Two dynamics make this exploitation of language possible:

- the regularization of language thanks to linguistic prostheses
- the banalization and manipulation of these prostheses (among which the autocompletion services) and their extension to all input interfaces.

4/ The commercially exploitable language is a predictable and regularized language

The commercially exploitable language is predictable by algorithms: the aim is to regularize language in order to predict it thanks to efficient models. The linguistic capitalism technologies thus lead to the language planning.

5/ Language is a capital

Language is a capital: this capital is transformed into a service of linguistic expression mediation that itself allows increasing the accumulation of capital. The objective of these actors is always to garner a larger capital thanks to predictable linguistic models. The growing linguistic capital is convertible into an economic capital, also growing.

II / Linguistic and cultural issues of the new technologic and economic context

-The role of pivot languages in automatic translators

Automatic translation algorithms are a precious service and allow a strong increase of the linguistic capital. However, this algorithmic intermediation in translation is not without retroactive effects on languages, cultures and idiomaticity.

Some results seem to be strange when one does not know how the system works: for example the Google translation of the French sentence “cette fille est jolie” [*this girl is pretty*] in Italian gives as a result: « [Questa ragazza è abbastanza](#) », literally this girl is “average” [*moyenne*]. Beauty has been lost in translation. This error is explained by the fact that the automatic translator goes through a pivot language, here English: “jolie” in French can be translated by “pretty” and “pretty” (which can mean “moyen” in French, i.e. average) is translated in Italian by “abbastanza”. It is indeed necessary, to produce an automatic translator, to have huge corpuses of identic texts translated from one language to another. Google being an American company, its tool has been built on pairs almost always associating English as pivot language: to go from French to Italian, it is thus required to go through an intermediary English translation.



-The cultural bias of algorithmic mediation: a new linguistic imperialism

The cultural bias of such a process is important. French and Italian are idiomatically relatively close languages, whereas English operates with a completely different logic and grammar: the passage through a pivot language leads to introduce in other languages **specific linguistic logics** and therefore **specific thinking modes**. At a global scale, a translation strings network is being set up and sometimes imposes, to translate an expression from a language to another one, to pivot through a series of intermediary languages.

The **linguistic imperialism** of English (or other languages that could play this role of pivot in other automatic translators) therefore has much more subtle effects than what suggest the approaches

which only study the “languages war”. Automatic translation also induces **linguistic and cultural effects** which are worth studying.



-Automatic description of images: an automatic generation of texts

The images automatic description systems are an example of texts automatically generated. By combining [neural networks](#) (which allow acknowledgment) with recurring networks, Google succeeded in creating systems able to **generate adequate descriptions of images**: if it is shown a big number of images and associated sentences, the algorithm allows **producing sentences correlated to the perceived visual elements**. To assess the quality of these expressions, an army of **human judges** is employed: the internet users assess the accuracy of sentences according to the described images. This systematic description of images by algorithm could be at the origin of new practices of indexation: today, it is still very difficult to find images from textual researches on the internet.

-The new creolization: hybridization between human writings and algorithmic texts

It is worth distinguishing within textual contents between:

-**primary resources** produced by **human** beings (scanned books, oral or written conversations)

-**secondary resources** produced by **algorithms** (automatically generated articles or translations).

However, it is difficult to build algorithms which are able to distinguish between these two kinds of sources. A growing number of texts are now produced by machines (it is the case of Wikipedia articles created by the Wikipedia bot, for example): as and when the quantity of secondary resources becomes significant compared to the one of primary resources, the statistic models are modified to integrate the algorithms vocabulary and turns of phrases. We are then proposed these ones in the form of suggestions or corrections by the autocompletion tools. Human and algorithmic texts are thus mixed to constitute new hybrid forms of writing, which the algorithms then analyze to structure our future linguistic productions. Language therefore integrates the linguistic bias of the machines and the constraints of global economy: a new retroaction of technology and economy on natural language.

-The transformation of writing practices by autocompletion tools

-[Finland recently proposed to give up the teaching of handwriting](#), to replace it by dactylography. The progressive abandon of cursive writing can lead to a loss of **orthography**. A citizen must have the competences that had the one who built the machine to be able to use it.

-Moreover, dactylography is no longer today the way to enter texts: the interfaces change very quickly, smartphones have led to a new form of interfaces and autocompletion has established itself.

-There is a fundamental difference between the practice of **writing** and the one of **autocompletion**. This difference is analogous to the one between tracing from memory a path on a paper sheet and to drive choosing directions suggested by a GPS map. When we write, we never are proposed other paths among which we should have to choose. During a conversation, persons complete their talk themselves. But with machines, their talk does not belong to them: the one who writes is no longer an **author**, but simply a driver who makes selections among predetermined speech paths. Writing becomes a selection of choice. With autocompletion, **writing = driving**.



-Autocompletion has powerful speed capabilities, but the ability to write implies a base: the risk then would be to write from a base [profiled on the user](#), which suggests him writing possibilities according to his past writings. Similarly, when we browse on websites adapted to our profile, we evolve in a world when no one meets: even if we believe reading the same web page, this one in fact is adapted to the user who consults it according to his previous uses. We only see in a loop preprogrammed information accurate to the interests of our profiles and calculated in advance by the means of algorithms.

-Problem of autocompletion today: the basis is secret. It should be mandatory for Google to publish its linguistic model: then it would be possible to build other linguistic bases in open access, and to obtain an experience of writing mediated by different bases of elocution and then a possibility of diversification.

Warren Sack – Out of bounds. Language limits, language planning, and linguistic capitalism



[Warren Sack](#) is a professor of “Film and Digital Media” at the University of Santa Cruz, a software developer and a specialist of algorithmic. He participated to the development of Software Studies and meditates on the political and social uses of new media: his researches focus on the theory and design of spaces for debates, on open-source software development and on analysis and conception of online educative environments. His articles “Images, nombre, langage, programme” and [« Une machine à raconter des histoires »](#) explore the relationships between computer programming and narrative writing.

Relying on Frédéric Kaplan’s lecture on linguistic capitalism, he displayed the issues of this thesis from the point of view of what is usually considered as “language limits” and shows its main consequences on contemporary semiotics and linguistics.

I/ Language limits and language planning

a) Linguistic capitalism and language limits

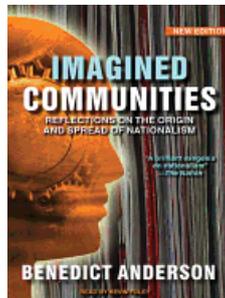
Warren Sack opened his lecture echoing Frédéric Kaplan’s linguistic capitalism thesis, which he finds really chocking. The economic model of Google indeed exceeds what is used to be considered as **language limits** – what the anthropology calls “[language ideologies](#)”, as defined by [Michael Silverstein](#) in 1979. The fact that not only novels, songs, poems and slogans but also words themselves can be sold, bringing \$ 5 billion per hour to Google, is deeply troubling. Warren Sack dedicated his lecture to reflect on these upheavals relying on the notion of language limits.

b) Definition of language limits and language planning

To explain this concept, he quotes the example of a joke by the comedian [Steven Wright](#). The joke is in two sentences: the first one is “I went to a restaurant that serves breakfast at any time” and the second one: “So I ordered a French toast during the Renaissance”. The second proposition, conflicting with the expectations aroused by the first one with their American way of life imagery, drives us to the heart of language limits. It is also the case of [Italo Calvino](#) when he apostrophizes the reader in the first paragraph of his novel [If on a winter's night a traveler](#). More generally, language limits are “composed of ideas we have about what language can do, what we do when we use language, and who can use language with whom and in what ways”. This notion involves numerous issues regarding social decorum and education. And the power to change language (that can have some writers and brands for example), or more generally **language planning**, is often perceived as beyond these limits.

II/ A parallelism between linguistic capitalism and print capitalism

a) “Print capitalism” (Benedict Anderson)



In the 19th century and at the time of [Noah Webster](#), the creation of references standards – such as dictionaries, thesauri or grammars – was considered as a crucial tool for the building of a **national unity**. “Regularizing the lexicon and the grammar of a language became a means to connect the population of a nation-state together – to keep them “on the same page” – through newspaper and other mass distribution publication of so-called “[print capitalism](#)””, Warren Sack explains. Here is the thesis of “print capitalism” as developed by political scientist [Benedict Anderson](#) in [Imagined Communities: reflections on the origin and spread of nationalism](#) (1983, French translation 1996). During centuries, language planning has been a political and religious instrument allowing the development of a sense of belonging, necessary for the national construction. But if there is any analogy between “print capitalism” and “linguistic capitalism”, then what is the aim of the latter, if it is not to build a nation? In **which contemporary political dynamic** can it be understood?

b) The break between linguistic capitalism and print capitalism

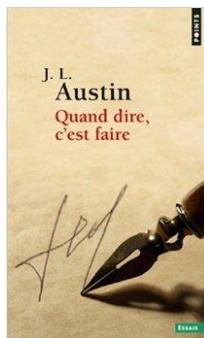
A noteworthy difference between print and linguistic capitalisms is that the “machinic assemblages” of the latter involve almost unimaginable issues of quickness and globality. Google’s era “imagined communities” differ a lot from the one described by Benedict Anderson, since they impose on us a writing and reading rhythm entirely different from the one of dictionaries and books. We are no longer at the age of journalism, characterized by a daily production, but in a situation where “each letter is scrutinized and coordinated globally in a fraction of a second”. As underlined by the Uzbek poet and journalist [Hamid Ismailov](#), [“With the blossom of social media in shape of Facebook and Twitter we are now talking not just about “journalism” but rather “houralism”, “minutealism” and “secondalism””](#).



III/ The relationship between language and the world in the age of Google

a) *New performativities of language*

In such a frame, the unavoidable question is to know how the contemporary language is connected to the world. At an age where it is not the geography but the **topology** of the world which is upset, and where information travels at the speed of light, it is worth wondering, from a **semiotic** point of view, what are the new connections between words and things. Focusing on the question from a historical perspective, one reminds that [Michel Foucault](#), in *Les Mots et les choses*, had shown that if the 16th century asked how a sign can designate what it means, the question of the 17th century was to know how a sign can be linked to what it means. But there is today a strong break with this so-called classical episteme and its interest in meaning that has spread in the 19th and 20th centuries. At the time of the web, the accurate paradigm to think the relationship between language and the world is rather the one of a **new performativity**, on the basis of [Austin's](#) thought but going much farther than what Austin could imagine. Texts and software make things with words, and they are means for instrumentation and manipulation rather than for representation. The performative powers of contemporary language, with their own quickness, are thus without precedent.



b) *Thinking the relationship between signifier and signified from C.S. Peirce's trichotomy*

Warren Sack bases this idea on the trichotomy developed by semiotician [Charles Sanders Peirce](#), who [distinguishes between icon, index and symbol](#). The **icon** is a sign that physically looks like what it stands for; whereas the **index** is correlated to what it stands for through a physical or chemical connection; and the **symbol** is linked to what it stands for through a psychological or cultural connection. But according to Warren Sack, this trichotomy should be completed by a **fourth term**, which would articulate the connection between a signifier and a signified by a computer network and would be a machine, or more precisely, a **calculation machine**. Just like a symbol, indeed, a machine is a sign that links signifier and signified through the cultural production of language planning. But unlike language yet, the machine has autonomous performative powers. This is why the equivalence between language and computational machines, already thought by [Alan Turing](#) in the 1950s, seems so strange to us. It seems to break with the fundamental architecture of our educative institutions, which have separated the arts of number ([quadrivium](#)) from the arts of language ([trivium](#)) during centuries. Therefore, imagining that language could be a kind of machine and reciprocally seems for

us to be out of the limits: it exceeds our common notion of **language limits** as conveyed by the pivot of cultural institutions.



Alan Turing

Conclusion

If the reality of linguistic capitalism is chocking, it is because it forces us to overcome the commonly presupposed language limits as elaborated by linguistics. In the [Chomskyan](#) reading of Descartes, machines cannot have the language, the use of the latter as an instrument of expression and creation being exclusively human. But Google engineers expressly conceive **machines that have the language**, even if it is a different kind of language. Consequently, linguistic capitalism only intensifies language commodification. It also exceeds the limits we assigned to it and the very definition of linguistics. The question that remains unresolved is the following: when the machines also have the language, who or what can be called human?