

Rafał Uzar

The Effect of Technology on Language and the Importance of Language Technologies

Kultura Popularna nr 4 (38), 30-40

2013

Artykuł został zdigitalizowany i opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Rafał Uzar

**The
Effect
of Technology
on Language
*and the Importance
of Language
Technologies***

There were still occasions when words printed on pieces of paper were the most convenient medium of communication.

(Clarke, A. C. 1987:126)

Rafał Uzar is a graduate of Lancaster University and the University of Łódź. He specializes in language technologies, corpus linguistics, translation, translation training and journalism.

1. Introduction

Language and technology have always been intrinsically connected. In 1960, Ted Nelson developed 'hypertext' as part of the Xanadu Project, which lay the foundations for the invention of the World Wide Web by Tim Berners-Lee almost thirty years later. Back in the 1960s, the celebrated writer, scientist and inventor Arthur C. Clarke envisaged a world in which computers could be accessed in one's own home and could provide us with information to help in our daily needs. Clarke talked about people being able to access their bank accounts and buy theatre tickets with a console the size of a book.

Floyd sometimes wondered if the Newspad, and the fantastic technology behind it, was the last word in man's quest for perfect communications ... It was hard to imagine how the system could be improved or made more convenient. But sooner or later, Floyd guessed, it would pass away, to be replaced by something as unimaginable as the Newspad itself would have been to Caxton or Gutenberg. There was another thought which a scanning of those tiny electronic headlines often invoked. The more wonderful the means of communication, the more trivial, tawdry, or depressing its contents seemed to be ... the newspapers of Utopia, he had long decided, would be terribly dull.

(Clarke, 1968:53)

The following paper is divided into three sections. The first section will touch upon the historical context of technological development with reference to language. The second section will focus on the idea that technology plays a leading role in changing language. This includes the internet as well as computers, tablets, smartphones, mobiles and the suchlike. The third section will deal with some of the language technologies and linguistic tools that are currently being used to manipulate, analyze and quantify language data, something never hitherto undertaken to this extent.

2. Typography as Energy

In parallel with the sci-fi prognostications and predictions of Arthur C. Clarke and William Gibson, the media theory scholar and philosopher Marshall McLuhan posited the idea of a "global village" in his 1962 *The Gutenberg Galaxy: The Making of Typographic Man*. Following in the footsteps of the great philosophers of human history, he also gave credence to the idea that human civilization had forever been riding on an evolutionary wave of intellectual development, spurred on by the invention of increasingly complex technologies.

In fact, McLuhan was convinced that contemporary human civilization had recently crossed an event horizon and was moving headlong into a new phase of history and growth. The future was tangible but, as always, a nebulous unknown.

The next medium, whatever it is (1) it may be the extension of consciousness, (2) will include television as its content, not as its environment, and (3) will transform television into an art form. A computer as (4) a research and communication instrument could (5) enhance retrieval, (6) obsolesce mass library organization, (7) retrieve the individual's encyclopedic function and flip into (8) a private line to (9) speedily tailored data of (10) a saleable kind. (McLuhan, 1964:10)

At a time where there were no PCs in our sense of the word, or the internet, both Arthur C. Clarke and Marshall McLuhan, to name but two scholars, were acutely aware of the feeling that the western world, and by extension human civilization, was at a remarkable tipping point and on the brink of technological progress. They both envisaged a world in which knowledge and data, embedded firmly in the linguistic sphere, would be of great if not central importance.

Popular science, to name but one area of study, has seen a slew of publications in recent years documenting the role that technology has had in mankind's development. A handful of example titles include: *The Big Ideas That Changed the World* (published 2010), *Inventors That Changed the World* (2011), *Breverton's Encyclopedia of Inventions: A Compendium of Technological Leaps, Groundbreaking Discoveries and Scientific Breakthroughs that Changed the World* (2012), *Inventions: A History of Key Inventions that Changed the World* (2012), *A History of the World in 100 Objects* (2012). In each publication, similar inventions can be found. The wheel, electricity, the airplane, the printing press, the telephone, the steam engine, the radio, the television, the automobile, the computer and the internet are all among the top technologies. Of the eleven mentioned here, at least five (plus computers) are intrinsically linked to linguistic communication. What is more, each of these new technologies has helped shrink our world.

And somewhere in the shadowy centuries that had gone before they had invented the most essential tool of all... They had learned to speak, and so had won their first great victory over Time... with the taming of fire, he had laid the foundations of technology... The tribe grew into the village, the village into the town. Speech became eternal, thanks to certain marks on stone and clay and papyrus. (Clarke, 1968:32)

These advances more often than not go hand in glove with language (or even literacy). The former have had, and will continue to have, an overwhelming effect on the latter, and in turn have a residual effect on our cognitive apparatus. Writing has given us the ability to perform complex calculations and contemplate abstract philosophical notions, therefore, giving us the weaponry for cognitive advancement.

Marshall McLuhan was aware of this fact and in *The Gutenberg Galaxy: The Making of Typographic Man* posited an alluring concept – human history can be categorized into four discrete phases of development:

Oral Culture
 Manuscript Culture
 Gutenberg Culture
 Electronic Culture

The catalyst for the shift from one ‘culture’ (viz. epoch) to another is the creation or invention of a new (communicative and/or linguistic) medium. Thus, the Manuscript Culture was preceded by the invention of (phonemic) writing systems, which also serves to explain why different societies, states, countries etc. can reside in different ‘cultures’ (to use McLuhan’s term). Therefore, the Sumerians and Egyptians were the contemporaneous initiators of what later became a Manuscript Culture thanks to the new ‘technology’ of cuneiform and hieroglyphic script, kick-starting the early phonemic systems of language recording of this new epoch.

As a side note and on a micro scale, it is interesting to consider Sumer (of the Sumerians) which was conquered by the Akkadian Empire whose language Akkadian was later displaced by Aramaic, the language of Christ. Many scholars, George Roux and John Sawyer to name but two, believe this latter displacement to be a direct result of the use of a “relatively easy alphabetic script” (Sawyer, J. 1999:14) by Aramaic rather than the “cumbersome cuneiform” (Roux, G. 1992:276) of Akkadian, the legacy of Sumer. Perhaps, this is one example of a technological medium having a direct impact on language.

Following the ‘cultural’ advancement of humanity through the invention of phonemic and alphabetic systems, the next thrust that lifted up the peoples of the Manuscript Culture was the technology of the moveable type in Johannes Gutenberg’s printing press which rocketed humanity to the giddy heights of the Gutenberg Culture.

In the same way that Gutenberg’s legacy empowered the peoples of its respective epoch, so Olivetti’s *Programma 101* initiated the Electronic Age in 1965. The *Programma 101* was the world’s first commercial personal desktop computer. However, does it end here? McLuhan’s four cultures could be augmented with a fifth (or perhaps an appendix to the fourth): the Information Age. Tim Berners-Lee’s invention not only poured oil on the flames of an already blazing technology, but it also democratized it, giving PCs the necessary software to ‘go forth and multiply’. The Information Age empowered the Electronic Age.

3. Technology as Catalyst

the emergence of knowledge societies makes literacy even more critical than in the past. Achieving widespread literacy can only happen in the context of building literate societies that encourage individuals to acquire and use their literacy skills.

(UNESCO Institute for Statistics 2008: 9)

A question that begs to be asked is whether the knowledge society/information age (or Information Age) that we live in now will bootstrap literacy. Will computers (and the internet) improve literacy or is full literacy needed to develop a knowledge society?

However, the big change that the internet has tangibly brought about is the democratization of language itself, allowing for its unfettered (and perhaps undisciplined) use. The internet has brought with it a new wave of language awareness. This meta-linguistic discourse is often manifested in the popular belief that the new medium has polluted or vulgarized the language. In the face of the liberalized use of language, a certain degree of conservative pushback is expected. Linguistic conservatives live in the fear that the language may be further corrupted which has led to calls for better writing, spelling and grammar:

pupils (or students as they are mysteriously called) are not taught such rules of spelling as may exist and certainly aren't tested on them. As for adverbs, subjects, objects or clauses, let alone such fabulous monsters as subjunctives, children are left in sublime ignorance of them. So [UK Minister of Education] Michael Gove's call for grammar to be taught in primary schools is sweet and catchy music to my ears.

(Wordsworth, *The Telegraph*, 6 July, 2012)

This is not an isolated opinion and in recent years there have been many calls for English schools to reinstall grammar, spelling, rhetoric, 'good' English as well as grammar schools back into the system, much to the chagrin of most educationalists (the United Kingdom Literacy Association, to name but one). They are fully aware that linguistic conservatism is nothing new and, when implemented, has generally not had a tangibly positive effect on the improvement of language skills amongst the youngest of learners. It is enough to turn back the clock exactly one hundred years to see that similar sentiments were also rife:

The English have no respect for their language, and will not teach their children to speak it. They spell it so abominably that no man can teach himself what it sounds like. It is impossible for an Englishman to open his mouth without making some other Englishman hate or despise him.

(Shaw, 1912:1)

In fact, we could turn the clock back another two hundred years, five hundred years or even a thousand years and read similar statements bemoaning the state of the language, be it German, French or even Latin. Every generation in nearly every state in the world has its fair share of linguistic warriors ready to fight the evils of linguistic modernity. Indeed, *tempus fugit*. And with it language also. Computers and the internet, therefore, should in no way be seen as 'harming' the language. In fact, literacy has improved over time, not the other way around.

New technologies are feeding into language and changing the way we use it. An obvious example of this is the use of text messaging on mobile phones, smartphones, tablets and the suchlike. To appreciate the impact 'texting' has

had on modern society, it is enough to look at the statistics regarding the use of SMSes in recent years.

Time Period	No. of texts
June 2002	1.38 billion
June 2004	2.11 billion
June 2006	3.51 billion
June 2008	6.31 billion

Table 1. Text Messages Sent in UK (Mobile Data Association)

Research figures from Deloitte tell us that 145 billion mobile text messages were sent in 2013. This figure is predicted to fall over the coming months. However, during the same period, 160 billion instant messages (IMs) were sent over mobile networks and this figure is set to rocket to 300 billion by the end of 2014. Despite another technological shift in the guise of IMs, to the detriment of SMSes, text messaging has already left its indelible mark on our language. SMS language has become as ubiquitous as it once was notorious and slated by linguistic conservatives. Words (and pragmatic markers) like OMG, LOL, CU have encroached into our everyday written and spoken language thanks to the medium.

Linguists have actually shown that when we're speaking casually in an unmonitored way, we tend to speak in word packets of maybe seven to 10 words... Speech is much looser. It's much more telegraphic. It's much less reflective, very different from writing. So we naturally tend to think, because we see language written so often, that that's what language is, but actually what language is, is speech. What texting is, despite the fact that it involves the brute mechanics of something that we call writing, is fingered speech ... Now we can write the way we talk ... But the fact of the matter is that what is going on is a kind of emergent complexity. That's what we're seeing in this fingered speech. . there is new structure coming up. (McWhorter, 2013)

Therefore, the medium of mobile phones has squeezed language into a new dialectal variation. This “emergent complexity” is a prime example (and the second in this paper after the example of writing systems) of how technology is in a very real sense fashioning the way we write and speak. This applies not only to texting but the internet in general. Words are also shifting in meaning thanks to new usages on the internet. *Liking* can now mean ‘befriending someone on a social networking website or approving of something on a social network’. *Unfriending* is often cause for concern and is the act of deleting a *friend* (!) from a social network.

This is not only limited to the English language. Journalist Aleksandra Krzyżaniak-Gumowska (*Newsweek Polska*, 2 August, 2013) lists a host of wonderful gems we can find in Polish. This emergent complexity has added to the vocabulary of modern Polish. This includes words and phrases like *slit focia* (eng. selfie, from ‘sweet photo’), *lajk* (form eng. ‘like’), the verb *hejtować* (to hate) or even *hashtagować* (to hashtag). Other social networking words like OMG (Oh my God), YOLO (you only live once) or SWAG (panache, style)

have crept into Polish so much so that they are an intrinsic part of the vocabulary of texting, instant messaging and the internet. Our third example of the technological medium affecting language.

The force with which we are witnessing language development as a result of technological progress is so powerful that even seemingly conservative circles are reluctantly having to accept the march of progress. Between 2013 and 2014, the *Oxford English Dictionary* (OED) added a host of new words to its pages. Examples include *interoperability* (ability to share information between systems), *bestie* (one's best friend – compare *selfie*, a photograph one has taken of oneself, typically with a smartphone and uploaded to a social media website), *to live-blog* (post commentaries while an event is taking place in the form of short blog updates), *TBH* (to be honest), *geekery* (obsessive interest in a specialist subject), and *Aspie* (a person with Asperger's syndrome). The fact that the last entry has entered the OED at this present time is of particular import as several scholars attest that individuals with levels of autism and Asperger's syndrome demonstrate competences that can be made use of in the computerized world, a very real example of how technology might be altering society (and perhaps another example of how technology is affecting change).

And the other thing is the high incidence of hackers like these who have characteristics which are consistent with Asperger's syndrome. Now I discussed this with Professor Simon Baron-Cohen who's the professor of developmental psychopathology at Cambridge. And he has done path-breaking work on autism and confirmed, also for the authorities here, that Gary McKinnon, who is wanted by the United States for hacking into the Pentagon, suffers from Asperger's and a secondary condition of depression. And Baron-Cohen explained that certain disabilities can manifest themselves in the hacking and computing world as tremendous skills, and that we should not be throwing in jail people who have such disabilities and skills... We need to engage and find ways of offering guidance to these young people, because they are a remarkable breed.

(Glenny, 2011)

4. (Language) Awareness as Liberation

As we have seen, technology is having a critical effect on language. However, we are only able to make sense of these changes thanks to the linguistic tools which have come to the fore in recent years allowing us to collect, analyze and quantify linguistic data. The rise of the personal computer has not only had a profound effect on language use, it has had a revolutionary impact on language research and given thousands of linguists access to myriad linguistic processing tools. The fact that the OED has added and continues to add such modern terms as those mentioned above to its dictionary is largely thanks to (personal) computers. Between 1993 and 2013 when John Simpson was

Chief Editor of the Oxford English Dictionary, a staggering 60,000 new words and meanings were added.

This does not necessarily mean that the English language has dramatically expanded in recent years (compared with previous times) but the advent of new language technologies has allowed linguists to catalogue in details how the language has changed. This has all been made possible largely thanks to the birth of corpus linguistics. The *corpus* is a large body of *real* language, a database or repository of linguistic data. On the face of it, this may not seem revolutionary as text-based research and studies were being undertaken in the past, for example by German stenographer Friedrich Wilhelm Kädin in 1897, who collected 11 million words of German in order to study the spelling conventions used in the language (McEnery & Wilson, 1996). However, the revolutionary aspect was applying the scientific method in linguistic research and making these large collections of linguistic data machine-readable, structured and sampled. This began in the 1960s with the work of Henry Kucera and W. Nelson, who developed the now legendary Brown Corpus and wrote the innovative *Computational Analysis of Present-Day American English* (1967), both of which paved the way for a seismic shift in modern linguistics from prescriptivism and to descriptivism.

Linguistics had hitherto been largely concerned with how the language *should* be spoken rather than how it *actually is* spoken. Although this is not to say descriptivism did not exist, it simply was not at the forefront of research. Linguists understood that empirical language research must lie with high-end statistical analyses. The key was to collect as much sampled linguistic data as possible in order to begin making informed suggestions about what was going on in language. As J.R. Firth's succinctly claimed: "You shall know a word by the company it keeps" (1957:11). This approach also led to other innovative computer corpus projects, such as the Lancaster-Oslo-Bergen Corpus (LOB) and the ground-breaking British National Corpus (BNC), which is seen as the benchmark for modern corpora today. It continues to be used by linguists, lexicographers, language teachers and sociologists alike for research on British English, British society and British culture.

Today, almost all corpora are 'tagged', that is they include part-of-speech data and sometimes other important information including semantic and pragmatic data (or even translation error data – see Uzar, R. 2006). The part-of-speech tagging of corpora is undertaken using automatic probabilistic systems to tag the words found therein (for more information see: Levin, Pezik, Uzar 2005; Uzar, 2006; Uzar, Waliński 2001). This gives linguists an incredibly powerful tool as, once annotated, a corpus can give us a snapshot of linguistic performance at a very specific point in time. The corpus, if sampled correctly, can give us insights that we might never have known before. This meta-knowledge is being utilized today in a host of different industries.

Modern lexicography is one area that has benefited immensely from the use of linguistic statistical work. *Frequency lists* tell us the frequency with which a word is used and *concordancers* show us the context (collocations) in which certain words and phrases are grouped together, inspired perhaps by Firth's ideas on context. To do similar meta-linguistic calculations in one's head would take days, months and perhaps years. This allows corpus linguists (lexicographers or sociologists) to track changes in the meanings of words over time or the appearance of new words or even the frequency of use of certain words over others and make inferences about society. Recent work by Tony McEnery of Lancaster University (McEnery, A and Xiao, Z. 2004; McEnery,

T. 2006) has given us insights about swearing in the English language with regards sex and class as well as fascinating studies of the attitudes of the modern English press towards Islam and the Muslim world.

Forensic linguistics is another field that has profited from these methodologies. Authorship tools and anti-plagiarism systems are based on corpus research. An example of this is research undertaken by Boston University, which found that large parts of Martin Luther King's 1955 doctoral dissertation had been plagiarized. Another example is recent work on discovering if Shakespeare was actually the author of his own works, as opposed to Christopher Marlowe, Francis Bacon or Edward de Vere (made famous by the film *Anonymous*). The work has been undertaken using comparative corpus methodologies. Forensic linguistics also makes use of corpus-based statistical techniques for, amongst others, intellectual property disputes and voice identification.

One of the most ubiquitous tools of the modern information age, *Google Translate*, is another that benefits from the new technology that is the corpus. It works by gathering vast collections of bilingual parallel translations (see: Uzar, 2006) and aligning them with each other, a task which is largely done automatically. This is followed by the extraction of key words, phrases and collocations which are retrieved as soon as the user needs a text translated. The quality of the translation largely depends on whether the *Google Translate* corpus has a text of similar style and register in its repository or not. The system learns by doing – the more texts it has and the more feedback it gets, the better it becomes. It is a self-learning system inspired by concepts of artificial intelligence. The once half-baked idea of a 'Universal Translator' may not be as ridiculous as it once was thought to be.

Another crucial language technology is Natural Language Processing – a marriage of computer science, linguistics and artificial intelligence. Developments in corpus linguistics owe a great deal to NLP. Synergistically, the construction of ever larger corpora hundreds of millions of words in size have allowed scientists to create better algorithms for statistical models that use probabilistic techniques (for example, part-of-speech tagging) which give us hope for artificial intelligence. Language research, therefore, is playing an important role in perhaps the most important technological step awaiting us. Speech recognition and speech production software are the tangible results of this drive towards AI.

British scientist Alan Turing is said to have been the father of artificial intelligence, and to some extent NLP. The Turing Test was a philosophical test designed to distinguish the behavior of a machine from a human. If a machine's behavior was indistinguishable from a human then could this lead to the supposition that the machine was intelligent, or even sentient? The idea might seem purely philosophical, however, the idea of the Turing test has been put to use with great success in the CAPTCHA system that most of us see on the internet during tests of identity. The CAPTCHA system (Completely Automated Public Turing test to tell Computers and Humans Apart) is a user identification procedure to determine whether a user is a human or not. It is in fact a reverse Turing test and is based on the human ability to recognize and decipher jumbled words.

5. Conclusion

Language technologies, although never often in the spotlight, are in fact a critical part of modern lives. Whether it is *hashtaging* a picture on *Instagram*,

translating a text on *Google Translate*, verifying the authorship of a text or simply using an online dictionary, society has become more meta-linguistically aware than ever before. In the Information Age all forms of data are invaluable. What is more, data is power. And linguistic data is perhaps the most valuable. Many of the world's national surveillance security systems seem to agree and make use of basic linguistic algorithms and concepts, including word and phrase frequency lists, key words or keyness (how crucial a word is to its context) in order to analyze, track and monitor the language we are all using. We are surrounded with examples of how language and technology are working together and developing each other.

Several examples of the synergetic and symbiotic development of language and technology have been mentioned here. Modern technology continues to shape language and society. Our first example was the invention of the cuneiform script by the Sumerians, which gave rise to a "Manuscript Culture" able to record language, thoughts and ideas in order to pass them on to the next generation. Another language technology, the printing press, led to another shift, to the "Gutenberg Culture". Computers have brought us into the "Electronic Age". Personal computing, it can be argued, is undoing the work of previous ages by doing away with the need for handwriting. Official letters and formal correspondence, if not typed and printed, are sent electronically nowadays. The need to write by hand is becoming increasingly rare. More recently, the "emergent complexity" of texting has demonstrated how the medium of mobile phones has created "fingered speech", a completely new variety of language. A *new* variety of language. Technology has had a profound effect on language: it is changing it, modifying it and augmenting it. What lies ahead can only be guessed at.

WORKS CITED

- Clarke, A. C. (1968) *2001: A Space Odyssey*. Hutchinson. London.
- Clarke, A. C. (1987) *2061: Odyssey Three*. Random House. New York.
- Gibson, W. (1984) *Neuromancer*. Harper Collins Publishers. London.
- Glenny, M. (2011) *Hire the Hackers!* TEDGlobal.
- Firth, J. R. (1957) *Papers in Linguistics 1934–1951*. Oxford University Press. London.
- Levin, E., P. Pęzik and R. Uzar (2005) Zastosowania baz danych w językoznawstwie [Using Databases in Linguistics] [in:] *Podstawy Językoznawstwa Korpusowego*. Łódź University Press. Łódź.
- McEnery, T. and A. Wilson (1996) *Corpus Linguistics*. Edinburgh University Press. Edinburgh.
- McEnery, A and Xiao, Z. (2004) Swearing in modern British English: The Case of Fuck in the BNC. "Language and Literature". 13:235–268. SAGE publications. London, Thousand Oaks, New Delhi.
- McEnery, T. (2006) *Swearing in English: Bad language, purity and power from 1586 to the present*. Routledge. London and New York.
- McLuhan, M. (1962) *The Gutenberg Galaxy: The Making of Typographic Man*. University of Toronto Press. Toronto.
- McLuhan, M. (1964) *Understanding Media: The Extensions of Man*. The MIT Press. London.
- McWhorter, J. (2013) *Txtng is killing language*. JK!!! TED2013.

- Oxford English Dictionary Online* (1989–2014). Oxford University Press.
- Orwell, G. (1949) *Nineteen Eighty-Four*. Random House. London.
- Roux, G. (1992) *Ancient Iraq*. Penguin. London.
- Sawyer, J. (1999) *Sacred Languages and Sacred Texts*. Routledge. London.
- Shaw, G. B. (1912) *Pygmalion*. Penguin Classics. London.
- UNESCO Institute for Statistics (2008) *International Literacy Statistics: A Review Of Concepts, Methodology and Current Data*. UNESCO Institute for Statistics, Montreal.
- Uzar, R. (2006) Korpusy w nauczaniu tłumaczenia i w pracy tłumacza [Corpora in translation training and the translation process], [in:] *Korpusy w Angielsko-Polskim Językoznawstwie Kontrastywnym*. Universitas, Kraków. (156–179)
- Uzar, R., J. Waliński (2001) Testing the Fluency of Translator, [in:] “International Journal of Corpus Linguistics”, 6 (155–166).