1 Mining the Cloud

Sofia Boschat-Thorez

Sofia Boschat-Thorez is a researcher, artist and educator whose work focuses on knowledge organization systems and the stories which can be extracted from them. She primarily investigates mundane objects such as archives, collections, datasets or museums, to recover traces of their creators ideas and circumstances (for better or worse). She also has an overlapping practice of archiving with an interest for variability, circulation, community and access. She is a member of Varia, a Rotterdam based initiative which aims at developing critical understandings of the technologies that surround us. She teaches at the Willem de Kooning Academy.
Varia: https://varia.zone

Sofia Boschat-Thorez
T.C.   Tangible-Cloud

(T.C.) Could you please tell us more about the start of the research Mining the Cloud you initiated together with Manetta Berends in 2016? Why did you get interested in the term cloud?

(S.B.T.) Prior to this specific research, Manetta and I had already spent a short year casually listing, discussing and collecting material on metaphors commonly used to speak about different aspects of the internet and its navigation. The documents we gathered included corporate speech, adverts, film extracts, patents, articles and more. We also made a database listing terms, source domains, target domains, aspects emphasized or hidden by the metaphor, texts, coinage date, author, and so on, provided we could find such information. We were trying to get an overview, as we believed it would help us grasp better what each was doing by contrast.

We then identified three categories for the metaphors we had : "raw material and transformation", "motion and experience", "local and global". Each of these underlined different aspects such as capitalist accumulation, time and geography. But we had to put the project on hold rather soon after we started because we were in the first year of our master and needed to graduate. When we picked this research up again, over a year had passed and we had decided to apply for funding. We had to reduce the scope of the research to make it more precise and feasible. When looking for the most relevant one, the "cloud" became an obvious candidate as it seemed to be reaching the end of its normalization process. It had just been on the news, in related literature, in adverts on the streets or elsewhere. Something critical seemed to had happened in front of our eyes and we wanted to make sense of it. Beyond that, we both shared a frustration with the difficulty to identify what the term was supposed to stand for, despite a good enough practical knowledge and understanding of how internet infrastructures work. What was meant by Microsoft praising "the cloud that is helping cure cancer" was quite different from Western Digital's "cloud of your own". An other term quite popular at the time was "mining". We were still in this "data is the new oil" discourse era, which coincided with the wider adoption of cloud computing infrastructures. So we thought it would be amusing to create a playful bridge through the means of pursuing research. And so this project became Mining the Cloud.

(T.C.) At some point, you decided to focus on the magazine Wired, and to "mine" all online issues in search for the term cloud. Could you please explain us how you proceeded and tell us about language processing tools you used? What problems did you run into when "mining" the general press? What is the magazine Wired, and why did you switch to it?

(S.B.T.) We made the choice to go for Wired very early in the research, at the time of the application. It first happened for practical reasons. As we wanted to retrace the emergence and spreading of the metaphor, we thought that news outlets would be a good place to look into. We initially wanted to avoid a US centric perspective so Manetta looked into Dutch newspapers and I did look at French ones. We did this manually to have a first sense of the material. It gave us very poor results. I remember getting mostly horse races results because of the Hippodrome de Saint-Cloud. I also noticed a tendency of journalists to attempt literal translations, making things even more incomprehensible. So we took the opposite direction. We decided to use a specialist North American source and contextualize it thoroughly. Both of us knew about Wired, which seemed to hold an aura of authority and legitimacy. Upon more research we got very excited because the magazine was founded in 1993, corresponding to the year the internet was opened to the public. It also held a close relationship with writer William Gibson, Nicholas Negroponte who founded the MIT Media Lab, Kevin Kelly and Stewart Brand from the Whole Earth Catalog, alongside other names familiar to us because of Fred Turner's famous book From Counterculture to Cyberculture[1]. With this book, Turner retraced the relationship between those San Francisco area tech entrepreneurs and the emergence of an ideology advocating computers as tools for individual liberation rather than military technology. Given the aura and importance of these protagonists in utopian and heroic narratives surrounding the beginnings of the internet, Wired seemed like an interesting fit.

At first we were quite concerned with the

legality of using this material. We tried to reach out to a few people at Wired to get an official permission but never managed to get any reply. Upon the suggestion of a member of the Dutch funding body we applied to, we eventually reached out to Bruce Sterling. We got a reply but his email was very obscure and unnecessarily dramatic, warning us about being jailed for hacking and things of that sort. And so we proceeded with the mining.

Technically, we got very lucky because their online archive had just been restructured.[2] All the content was accessible through years, and the URLs were very cleanly organized. This had proved very useful for crawling through the website's content. We wrote a script using Python, for which we used a web-crawling framework called Scrapy. We retrieved around 180,000 articles, covering twenty-four years of content. Then, we made an other python script using the libraries HTML Parser and NLTK (Natural Language ToolKit) to retrieve extracts from the articles which contained the term cloud and the short paragraph in which they were used. All the found metaphors were saved in a big JSON file, including the year, title and URL of the article, alongside the context. Overall, the process was extremely smooth.

(T.C.) What were the limits you felt to these language processing tools?

(S.B.T.) We ended up using more manual labor than language processing tools. Through the script, we managed to retrieve all instances of the word cloud, and the few sentences around it, but we didn't have the capacity to make sense of whether the use was metaphorical or not. And actually we came to realize that, even as a human reader, making sense of whether a term is used metaphorically or not can be complicated. Sometimes the word cloud is used to simply mean a meteorological phenomenon, but in other instantiations it can be a shape, an allegory in which a cloud can really be a cloud, a product, a brand name or a technological component. And what the term stands for is not always very obvious.

In order to annotate and analyze the metaphors, we tried to come up with a method. We reached out to trained linguists and attended the Metaphor Lab Amsterdam's winter school at the University of Amsterdam. But we found it quite difficult to connect our very specific research with a method that was focused on metaphor recognition in general, rather than on subtle differences in meaning between different instantiations of the very same metaphor. We decided to be more specific on that last point and read A Prehistory of the Cloud by Tung-Hui Hu.[3] It allowed us to identify different areas covered by the use of the metaphor : network, virtualization, data storage or data mining. So we decided to manually attempt to identify whether, from the context, we understood that the term was used as a metaphor to cover one or several of these aspects. Because it was sometimes tricky, we eventually included a field in the JSON in which we detailed what we understood. Still, describing precisely what is meant by the use of a metaphor remained extremely complicated.

(T.C.) What conclusions, or interesting aspects did you draw out of this research? Did it help you grasp the term cloud better?

(S.B.T.) Throughout our research, we managed to look at the usage of the term from 1993 until 2017, which allowed us to have a sense of the evolution of the discourses in which it appeared. An interesting thing was that, alongside non metaphorical uses or far fetched cyberpunk fantasies, we found some relevant uses of the term from very early. Bruce Sterling spoke of "a whole cloud of hangers-on, suppliers, dealers, niche marketeers, and brand-new startups"[4] in 1993, and Kevin Kelly mentioned "the Internet Cloud"[5] in 1994. Conceptually, the idea of cloud computing didn't seem new at all. An other important moment for us was when we realized that many companies and products mentioned had the word cloud in their name, something that we noticed again was already happening in the '90s. So we made more research toward this direction and saw a clear the connection between branding and the emergence of this word in the context of computing. The first mention we could find of the term cloud computing was in an internal document at Compaq from 1996. Ten years later, with Amazon's Elastic Compute Cloud service, the term started to be more frequently used by big tech companies. In

2007, Dell tried to trademark the term cloud computing but the US Patent and Trademark Office eventually refused it.

Getting back to the material we used for this research—which was Wired magazine—it made perfect sense to start seeing things under this lens. Ultimately, the magazine's main readership was initially people in managerial positions (84% of the readers in the very beginning[6]). Wired was geographically, socially and ideologically in close proximity to Silicon Valley entrepreneurship. So the articles reflected and spread libertarian ideas on technology, alongside the vocabulary to be adopted by companies in the sector. There seemed to be a clear relationship between Wired's writers "predictions" of the future and the products and branding concepts then created by its readership. Many of the quite recently adopted software products and technologies were already present somewhere in corporate imagination for some time. There is an interesting series of adverts from 1993, directed by David Fincher for AT&T, which foresaw a lot of the ways in which technology is currently implemented. It mentioned the possibilities of borrowing books from far away, attending meetings bare feet, watching a movie the minute after wanting to do so, sending faxes from the beach, and the list goes on. Rather than "accurate predictions", ideas of the future were produced by people deemed as legitimate and these ideas spread and set the direction for companies to make it happen.

Ultimately, the research didn't help us come up with a definition of the word itself. In 1999, William Gibson declared that the cloud's usefulness lies in its vagueness, and we believe it is most likely the reason for its widespread use. The term appears to work alike the representation of clouds in patent drawings, referring to necessary parts of an infrastructure one doesn't wish to describe and make visible. The issue with this is that, while the audience for patent drawings might be capable to fill the gaps, the potential users of cloud computing services (literally anyone) may not be experts in the field. I remember us wondering what could happen if we departed from descriptions of different cloud computing elements and came up with new metaphors. Could there be other terms that would rather help us grasp and perhaps differentiate all the things laying behind the cloud? If the vocabulary was coined by users on the basis of how they relate technology to familiar concepts, and if this vocabulary was intended to facilitate understanding among non-experts, what would these metaphors look like?

(T.C.) You said you had to give up on that project. Could you please tell us why, and talk about artistic research and its context? How do you regard this research now, six years later?

(S.B.T.) This project was our first project after graduation, and so our first funding application. Although we wanted to make something quite large and focused, we started with less money than we had initially applied for. This prompted us to reduce the scope of our work, but also to individually seek other sources of income. It was also quite hard at one stage to make time for it, not just physically but also to be mentally present into a project when the resulting work pace was so fragmented. Keeping the rhythm under these "project based" frameworks for artistic research is a bit hard sometimes. So as usual, the material conditions under which one can do research are crucial in what can be delivered. It leaves little room for long term development and puts pressure on production. Looking back, it feels a bit like it could have been the beginning of a bigger project. The most interesting part was the moment when we tried to make sense of the metaphors and explain in our own words what we thought we recognized, or not. We even attempted to make schemas to map visually the "clouds" referred to in the articles we were reading. Had we had more time, money, and perhaps motivation, it could have been interesting to spend more time on that specific exercise and open the process. It could have generated some great material to work further with.

[1] Turner, Fred. From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism. University of Chicago Press, 2008

[2] Chirlin, Benjamin. "How We Moved 34,000 WIRED

Pages to One Site in 9 Hours." WIRED, 30 Oct. 2015, https://www.wired.com/2015/10/cyphon-wired-archive-migration/

[3] Hu, Tung-Hui. A Prehistory of the Cloud. MIT Press, 2015.

[4] Sterling, Bruce. "War Is Virtual Hell." WIRED, 1 Jan. 1993, www.wired.com/1993/01/virthell.

[5] Kelly, Kevin. "The Electric Paperless Prototype." WIRED, 1 June 1994, www.wired.com/1994/06/paperless.

[6] "The Killer App." The Baffler, 10 June 2020, thebaffler.com/salvos/the-killer-app.

Dodge, Martin 2008. Imagining Internet Infrastructures: Spatial Metaphors and Scientific Inscription. (Ph.D. thesis). Manchester University, Manchester.

Hu, Tung-Hui 2016. A prehistory of the cloud. (First MIT Press paperback edition ed.). Cambridge, Massachusetts: The MIT Press. (OCLC: 962421817)