Tommaso Venturini, Daniele Guido

Once Upon a Text: an ANT Tale in Text Analysis

(doi: 10.2383/72700)

Sociologica (ISSN 1971-8853)
Fascicolo 3, settembre-dicembre 2012

Ente di afferenza:
()

Once Upon a Text: an ANT Tale in Text Analysis

by Tommaso Venturini *and* Daniele Guido

10.2383/72700

The Problem with ANT

The founders of Actor-Network Theory have never been happy with the name given to their approach. In a symposium dedicated to the future of ANT and transcribed in the book *Actor Network Theory and After*, ANT founders vie in criticizing the expression "Actor-Network Theory":

> John Law: "The desire to know clearly what we are talking about, the desire to point and name, to turn what we now call ANT into a 'theory', I believe that all of these things have done harm as well as good. 'Have theory, will travel'" [Law and Hassard 1999, 7].
> Bruno Latour: "There are four things that do not work with actor-network theory; the word actor, the word network, the word theory and the hyphen! Four nails in the coffin The third nail in the coffin is the word theory […] Far from being a theory of the social […] it always was, and this from its very inception, a very crude method to learn from the actors without imposing on them an *a priori* definition of their world-building capacities" [*ibidem*, 15, 19-20).
> Michel Callon: "ANT is not a theory. It is this that gives it both its strength and its adaptability. Moreover, we never claimed to create a theory. In ANT the T is too much ('*de trop*'). It is a gift from our colleagues. We have to be wary of this type of consecration especially when it is the work of our best friends. *Timeo danaos et dona ferentes*: I fear our colleagues and their fascination for theory" [*ibidem*, 194].

Manifestly, what bugs Law, Latour and Callon the most is the word *theory*. In one way or the other, they all reject the label of *theorist* and swear that their intention

has always been to liberate social research from the burden of social theories. Encouraging scholars to learn from social actors more than from sociology handbooks, ANT shares many resemblances with ethnomethodology [Garfinkel 1967] and grounded theory [Glaser and Strauss 1965; Glaser and Strauss 1967]. It does not flee from generalization, but claims that speculations must follow data and not the other way around. Dump all structuralist assumptions, teaches ANT, there is nothing behind or beneath collective phenomena but the trajectories of the actors and the networks that they build. Follow the actors, deploy their networks, and that's all you need to know about sociology.

Ironically though, while the notion of actor-network has gradually been accepted as a theoretical description of collective dynamics, ANT is still criticized for providing little help with operationalization. In *On Difficulty of Being an ANT*, Bruno Latour acknowledges the problem. Dialoging with a student eager to *apply* his theory, he demolishes all received ideas on ANT's practice and cheerfully concludes:

> S — I can't imagine one single topic to which ANT would apply!!
> P — Beautiful, you are so right, that's exactly what I think…
> S — That was not meant as a compliment.
> P — But I take it as a true one! An application of anything is as rare as a good text of social science.
> S — May I politely remark that, for all your exceedingly subtle philosophy of science, you have yet to tell me how to write one... [Latour 2005, 151].

Having long worked with Latour, all my sympathies go to the student. The problem with ANT is the gap between the originality of its theoretical premises and the practical tools available to materialize them. Most ANT studies rely, in fact, on a conventional ethnographical approach. From its beginnings in an anthropology of the laboratory [Latour and Woolgar 1979] to the "technology turn" [Bijker and Law 1992], to the most recent works on law [Latour, 2002], medicine [Mol 2001] and economics [Callon, Millo, and Muniesa 2007], actor-network theorists have kept working with the usual toolkit of observation and interviews.

To be sure, there is nothing wrong with ethnography, except that its underlying tenets are hardly compatible with the ANT approach. Qualitative research builds on the idea that interesting things happens *locally* and can be observed *directly*. But if the *-network* in "actor-network" means something, it is that there is no such a thing as a *local interaction*. Interactions are always at the crossroad of trajectories that begin and arrive elsewhere. To trust qualitative methods, one has to believe that some distinction can be drawn between action and context, between what happens *here and now* and the infinite series of influences that frame this moment. Unfortunately, this is precisely the point that ANT has always challenged.

At the same time though, ANT never liked quantitative methods and their rough way of ironing over the folds of collective existence. Favoring exceptions over norms, controversies over consensus, change over stability, actor-network theorists have always been wary of averages and aggregation. If the *actor-* in "actor-network" means something, it is that in social sciences there is no law of large numbers: collective dynamics are constantly and substantially deviated by the trajectory of every single agent involved. Far from being the opposite of "individual," "collective" means that plenty of individualities are to be taken into account.

The Quest for Quali-quantitative Methods

Longing for an empirical grasp, but dissatisfied with both qualitative and quantitative approaches, actor-network theorists searched for the grail of social sciences: the quali-quantitative methods. In their quest, ANT scholars soon encountered network analysis and immediately recognized an elective affinity with a method that allows exploring millions of associations while keeping trace of each single data point.

More precisely, it was scientometrics that stirred the attention of ANT founders. Not only because scientific production had always been a privileged subject for ANT, but also because the alternative tradition of "social network analysis" seemed too human-centric for the champions of the generalized symmetry principle. Scientometrics with its capacity to trace citations and scientists, keywords and institutions and to follow them across oceans and disciplines seemed the perfect materialization of ANT.

Scientometrics, however, has a major disadvantage: drawing on a strict definition of what counts as an author, a paper, a citation, it is difficult to apply outside the formalization of scientific literature. Striving to extend scientometrics, ANT scholars moved from citation analysis to co-occurrence analysis [Callon, Law, and Rip 1986, Teil and Latour 1995] and helped develop several pieces of software (Leximappe, Candide, Calliope, Réseau-Lu, Prospéro). It was through digital cartography, however, that ANT moved definitively beyond the limits of scientific literature.

The first trace of the encounter between ANT and digital methods is to be found in 1998, in the occasion of the first conference of the *Virtual Society?* program [Woolgar 2002]. In his speech, Bruno Latour [1998] presented for the first time the idea that digital traces could provide the materialization of interactions that ANT was looking for:

> Once you can get information as bores, bytes, modem, sockets, cables and so on, you have actually a more material way of looking at what happens in Society. Virtual Society thus, is not a thing of the future, it's the materialisation, the traceability of

> Society. It renders visible because of the obsessive necessity of materialising information into cables, into data.

In the audience of the conference was a young American sociologist, Richard Rogers, who, in the following years, developed a countless series of tools and methods to put digital traces at the service of social sciences [see Rogers 2005; Rogers 2009][1]. The most famous of these tools, the IssueCrawler was explicitly developed to materialize ANT:

> Bruno Latour (1998), argued that the Web is mainly of importance to social science insofar as it makes possible new types of descriptions of social life. According to Latour, the social integration of the Web constitutes an event for social science because the social link becomes traceable in this medium. Thus, social relations are established in a tangible form as a material network connection. We take Latour's claim of the tangibility of the social as a point of departure in our search [Rogers and Marres 2002, 342].

The IssueCrawler is a simple piece of software that follows and traces online connections around any issue of public debate [Marres and Rogers 2005]. Drawing on the standardization of reference provided by web protocols, the IssueCrawler extends the methods devised for scientific citations to hyperlinks with a twofold gain: allowing scholars to follow actor-networks beyond scientific literature and cutting the cost of tracking social connections.

The Médialab Toolkit

Encouraged by the success of Rogers' *Digital Methods Initative*, Bruno Latour created the Sciences Po médialab with the explicit mission to harness digital traceability and materialize the dream of quali-quantative methods [Venturini and Latour 2010].

Not surprisingly, network analysis (Gephi and Tubemynet) plays a pivotal role in the médialab toolkit, together with scientometrics (Scholarscape and Sciscape) and web cartography (IssueCrawler and NaviCrawler). The techniques originally identified by ANT scholars still provide the soundest methods to operationalize actor-network theory and, accordingly, they were the first to be offered by Sciences Po médialab.

[1] See also www.digitalmethods.net.

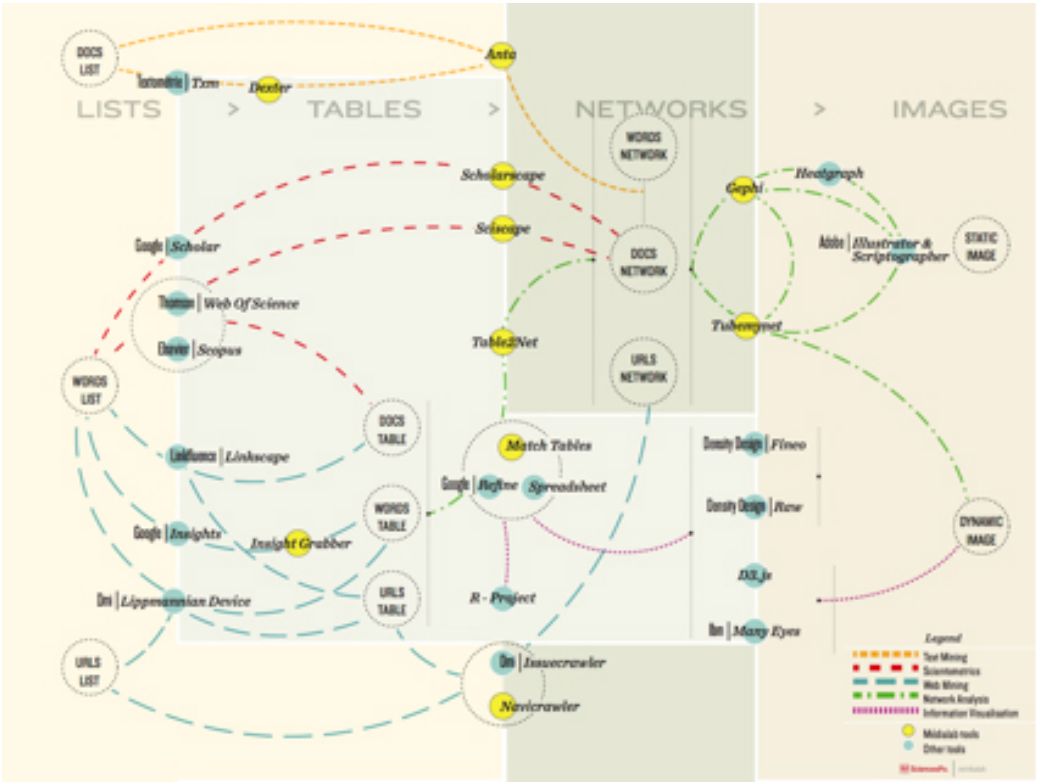

FIG. 1. The médialab toolkit.

Since the médialab opening in May 2009, we have received dozens of researchers searching for an empirical grasp on ANT or interested in experimenting with quali-quantative methods. Most of them are fascinated by the network analysis tools (Gephi, gephi.org, in particular) and the way these tools turn aggregation and disaggregation into a seamless navigational experience [Bastian, Heymann, and Jacomy 2010]. Zooming from millions of connections to a single remarkable data point (and back), researchers feel that the quali-quantitative methods are at last at hand [Venturini 2012]. Their enthusiasm, however, cools down when they realize that all the maps we show them are limited to scientific or online discourses. Sciences studies scholars appreciate scientometrics and media studies scholars love web cartography. But apart from them, nine tenths of the researchers we met were bitterly disappointed.

Researchers can arrive to quali-quantitative methods from either side of the great methodological divide and their difficulties vary accordingly. Those who come from quantitative methods have little problem with data: most of their tables and databases can easily be transformed into networks. To such end, we even developed a little tool, Tablet2Net, capable of turning a table into a network in a few click. To be sure, this is not enough to move from quantitative to quali-quantitative. Most of the time, the data available are too discontinuous to allow a seamless zooming in and out,

and researchers are too used to confirmatory data analysis to enjoy the exploration introduced by quali-quantitative techniques [Tukey 1977].

Qualitative researchers, on the contrary, arrive at the médialab bringing rich data and longing to explore them. Their problem is that qualitative data cannot be easily fed into network analysis tools. Quantitative data can have many different forms (from a video recording to the very memory of the researcher), but they are often stored in a textual format (i.e. interviews transcriptions, field notes or archive documents…). The question therefore becomes: how can texts be explored quail-quantitatively? Or, more pragmatically, how can texts be turned into networks?

Introducing ANTA

ANTA or Actor-Network Text Analyzer has been developed by the médialab with a straightforward objective: to transform a set of texts in a network. This is, of course, easier said then done and developing the current version of ANTA took two years of trials and errors (mostly errors). Groping for ANTA, we were spoiled with choice. From a conceptual point of view, there is a multitude of textual elements that can be turned into nodes (the texts themselves, the paragraphs, the sentences, the phrases, the words, the lemmas…) and a multitude of elements that can be turned into links (the sharing of lemmas, exact words, group of words, arguments, items of style…). From a technical point of view, digital text analysis is as old as computers and countless scripts and algorithms are available at the plainest web search. Computational linguistics and natural language processing made great strides in the last few years as sophisticated methods as sentiment and latent semantic analysis showed.

Coming from scientometrics and web cartography, we were reluctant to move from explicit connections (scientific citations and hyperlinks) to implicit ones such as the sharing of arguments. Therefore we took a radical decision: our tool would have to be simple and stick to the KISS principle. We *Kept It Simple (and) Stupid* because we wanted to privilege interpretability over everything else. We wanted researchers to be able to read straightforwardly the graphs we handed them and know exactly what is in them. Most of ANTA's features are derived from this engagement.

The first distinctive feature of ANTA is that it only recognizes two types of elements: documents and expressions. This choice is less obvious than it seems, in particular as concerns documents. Though documents are the usual form in which texts are given, they tend to be disregarded by text analysis tools. To be sure, there are good reasons not to trust documents. For one thing, they may be incredibly diverse: an haiku of seventeen syllables is a text exactly as a novel of thousands pages.

A document can be structured or unstructured; it can be linear as in a prose or non-linear as in glossary; it can mix languages and includes "special elements" such as indexes, titles, bibliographies and so on. That is why most tools prefer neglecting text and concentrating on more precisely defined elements such as phrases, paragraphs or words' neighborhoods. Relying on documents, however, has the advantage of drawing on the "natural" boundaries of the corpus. It may be difficult to explain how phrases or paragraphs have been delimited. It is much easier for documents, as their unity is granted by the choice of the corpus construction. Focusing on documents has another advantage: documents are the protagonist of the enunciation. Even when they cannot be attributed to a unique and clearly identified author, they remain there to support the speech act. In the world of ANTA, a graph of documents is a close approximation to an actor-network.

As for expressions, they are simply defined as words or groups of words. For the sake of simplicity, we decided to renounce to lemmatization and recognition of inflected forms. ANT has taught us that, in the construction of collective existence, the devil is in the details. Most of the times the conjugation of verbs and the declension on nouns does not make a big difference. Sometime it does. Fail to distinguish "we are changing the climate" from "we may be changing the climate" or "the cause of climate change" from "the cause*s* of climate change" and you will lose the difference between climate scientists and climate skeptics. This is why, instead of automatic lemmatization we prefer to give researchers the possibility to merge equivalent expressions (more on this in the next paragraph).

Although ANTA does not lemmatize, it is capable of recognizing groups of words regularly occurring together or *n-grams*.[2] In the current version, expressions are extracted through AlchemyAPI (www.alchemyapi.com), a free online service providing n-grams extraction in eight different languages. Though AlchemyAPI offers a trustworthy service, we don't like relying on it. In particular, we don't like that the service is offered as a "black box" and that the exact extraction algorithm is secret. Something that is perfectly reasonable from a commercial viewpoint may be a problem for research, and this why we are working to implement an open-source version of the n-gram extraction. We want to be able to explain precisely how expressions are identified (out of the multitude of n-grams present in a text). There is, however, something nice about AlchemyAPI: its extraction algorithm has been developed for "entity extraction," which means that it favors that expressions that refers to "real world entities" such as persons, institutions, companies, organization,

[2] In computational linguistics, an n-gram is defined as a contiguous sequence of *n* elements in a text. The elements composing an n-gram can be letters, syllables or words (as in the case of ANTA).

geographic features, cities, countries and so on. This is a feature that we would like to preserve (by drawing on open projects such as Freebase or BDpedia) for at last two reasons. The first is that "named entities" are generally less ambiguous than other types of expressions. The second and more important is that named-entities provide a good operationalization of a crucial category of actors involved in enunciation. We already discussed the role of documents of protagonists of the textual enunciation. Yet, as the semiotics of *débrayage* explained [Greimas and Cortés 1979], any speech act mobilizes two sets of protagonists: those who say something (the documents, in our case) and those who are said something about (the named-entities, in our case). In textual analysis, a bipartite graph of documents and named-entities constitutes the closest approximation to an actor-network.

The second distinctive feature of ANTA is that it only considers the simplest type of connection between documents and expressions: plain occurrence (an expression is connected to a document if it is contained in it). The reasons why we preferred occurrence over more complicated type of relations is that it keeps documents in the foreground. Even simple co-occurrence (connecting two expressions as they appear together) seemed inappropriate for it would have made documents disappear. Of course, it is possible to transform a bipartite network of documents and expressions[3] into a mono-partite network of documents or expressions (and the next version of our tool will allow such transformation). The default ANTA export, however, will remain the documents-expressions graph that directly displays the most basic information: which expressions are used by which text.

A third distinctive feature of ANTA is that all the efforts saved by *not* implementing complex linguistic methods have been invested in the design of the user-interface. Deciding to avoid sophisticated linguistic tricks, we knew that this meant moving the burden of analysis on the users' shoulders. We already explained how renouncing to automatic lemmatization increases the need for manual expression merging, but this is not the only example. In general, in developing ANTA we automated as little as possible the work of text analysis and we preferred equipping the users with an ergonomic interface allowing them to perform their investigation quickly and with the less possible effort.

The best example of this *user-intensive approach* to text analysis is the way ANTA implements the selection of expressions. Choosing a limited number of expressions for the analysis is a crucial operation of any text-based research and for a simple

[3] In the jargon of network analysis, a bipartite graph is a network composed by two different types of nodes, such that two nodes of the same type are never directly connected. The networks produced by ANTA are bipartite because their edges always connect an expression and a document (and never two documents of two expressions).

reason: even relatively small corpuses are composed of too many expressions to follow them all. Below we are providing a little example using eight tales from the Grimm brothers' repertoire. They are all short stories (Snow White, the longest of our corpus, is only five pages long) and yet ANTA extracted more than six hundreds expressions from them. In public presentations of ANTA we are often asked if AlchemyAPI is capable of recognizing all the n-grams occurring in a text. The answer is "no, and thank God!" AlchemyAPI only recognizes the most frequent and important expressions (remember that it is originally intended for named-entity extraction).[4] And yet, even an average corpus of a few hundreds pages, can contain dozens of thousands of expressions. Selecting a few hundreds of them is a necessary step of the analysis. In most other text analysis tools such step is heavily automated and the expressions are selected on the basis of their frequency, their specificity or through the use of a predefined dictionary. In ANTA, we propose two metrics to filter the expressions (the number of documents in which they occur and the occurrence frequency in the document where they occur the most, normalized by the size of the document), but we leave to the user to decide where to cut. Also, ANTA encourages researches to be thrifty in the use of the automatic filtering and provides a selection interface that has been carefully designed to speed up the manual filtering. As we will show in the example below, the same user-intensive approach has also be applied to the categorization of expressions and documents.

ANTA, the Drive-through

To illustrate the functioning of ANTA, we will now carry out a quick analysis of eight well-known tales from the *Household Stories* of the brothers Grimm. A typical ANTA projects goes through five main phases that roughly corresponds to the five main pages of the software interface.

The first operation when working with ANTA is to upload the documents constituting the corpus (see figure 2). In the example, every document is a txt file containing one of the chosen tales.

The second operation consists in assigning metadata and tag to each document of the corpus. As for the metadata, in order to facilitate the expression extraction, researchers are asked to identify the language used in each of the document of the corpus. Optionally, researchers can also timestamp documents (usual-

[4] AlchemyAPI being a black box, we don't know exactly how the expression extraction is performed. We know, however, that AlchemyAPI only extracts n-grams of more than one word and less than four (unless it recognizes them as named entities).

ly using their date of publication) in order to follow the temporal evolution of the corpus.

As for the categorization, researchers can apply to documents any classification filling the research objectives. Common categories include author, type, and subject, but more specialized qualifications can be used. For instance, in our Grimm stories, we could have categorized documents according to the nature of their protagonists (human/not-human) or their gender (see figure 3). Categorizing documents is not necessary, but it can be extremely useful to provide answer to questions like: do different categories share the same language or are they characterized by specific linguistic expressions? This categorizing process has been conceived to be as flexible as possible: researchers can always modify the tags and the metadata they used as their research interests evolve over the time.

As soon as researchers upload their documents ANTA starts exchanging with AlchemyAPI performing the expression extraction. Depending on the size of the corpus, such operation can take a few minutes or a few hours, but ANTA handles it in the background and "server-side."

Once the extraction is completed, researchers are asked to perform the third operation in ANTA's workflow: the preliminary semi-automatic filter of the expressions. All the extracted expressions are represented in a scatter plot graph having the document frequency in abscissa and the maximum term frequency value in ordinate. This graph provides researchers with an overview of their corpus and allows them to establish a minimum and maximum value on each axis, thereby regulating the number of entities to include in the analysis. In most corpora, eliminating the expressions appearing in only one document excludes 90% of the extracted expressions (see figure 4). We recommend researchers to be very cautious in the use of the automatic filter and to prefer (within the limits of their time constraints) the manual filtering.

The forth step is the longest but also the most important. It consists of three manual operations usually performed together: the filtering, merging and tagging of the expressions (see figure 5).

FIG. 2. "Include document" section in ANTA. Tags and language can be assigned while uploading texts.

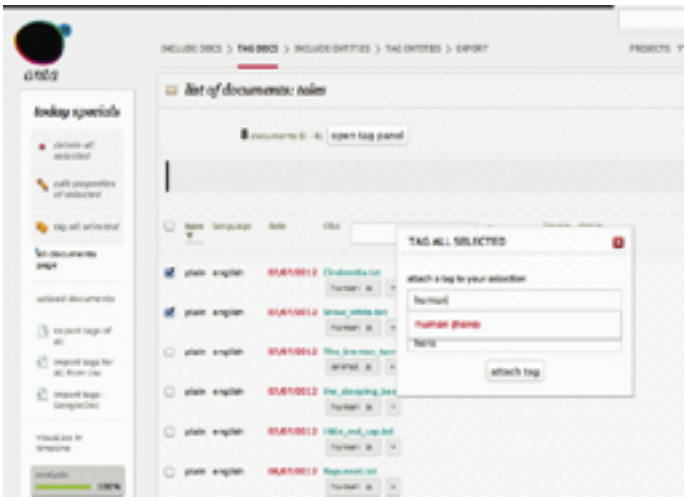

FIG. 3. An overview of the "tag documents" section. Documents are listed along with their tags and a series of functionalities are available

FIG. 4. "Include entities" section provides an overview of the corpus composition and a "crop" function that adjust the number of expressions included in the network.

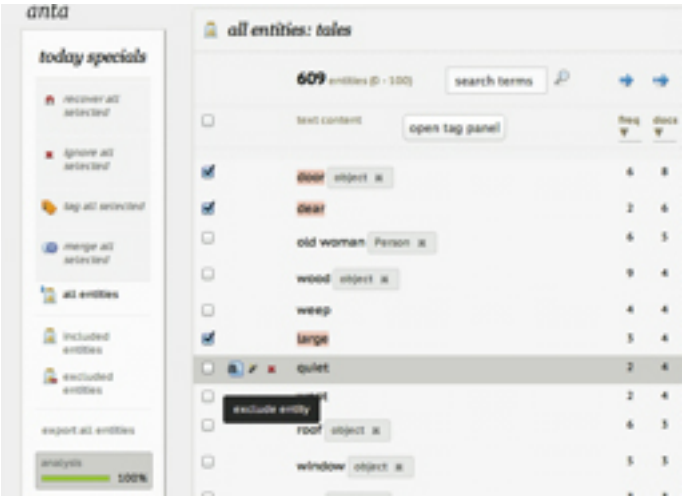

FIG. 5. The table of extracted entities.

The manual filtering is crucial because it allows entities to be reduced to a set size appropriate for analysis, but also recovering important entities that could have been excluded by the automatic filtering. In general, we recommend researchers to exclude:

a) Errors and expressions that make no sense (i.e. "to-morrow," "looking-glass speak," incorrectly extracted from the sentence "When she heard the looking-glass speak thus she trembled and shook with anger").

b) Expressions directly related to the type and therefore common to all the documents in the corpus (i.e. "Once upon a time").

c) Too generic expressions (i.e. "dear," "beautiful," "iron").

d) The expressions that would need require context to be interpreted (i.e. "little bit," "long way").

Besides including and excluding, researchers are also meant to check and refine the extraction performed by AlchemyAPI, correcting errors and merging equivalent expressions. This is the most delicate part of the work with ANTA and the one in which the expertise of the researcher is most precious: who else can decide that "poisoned apple" and "poisonous apple" are synonyms in Snow White, while "the apple" in Hansel and Gretel refers to a different item? In order to facilitate the researchers' work ANTA proposes two powerful search functions: the possibility to see the expression in the verbatim of all the sentences in which they appears and the possibility to look for "similar" expressions. The latter is especially convenient because AlchemyAPI returns "the apples" and "apples," for example, as distinct entities. If merging can reduce the number of entities and thereby facilitates the interpretation of the graph, the choice of a specific expression can be crucial to distinguish the discourse of a given document ("little girl" and "little maiden" for example may be used to distinguish girls of different age).

As for the tagging of expressions, ANTA proposes a preliminary one based on AlchemyAPI named-entities extraction engine. AlchemyAPI knows, for example, that "Rosamond" is a person name, that "Germany" is a country, and "Bremen" a city. Such tags are therefore automatically attached to related expressions, but mistakes are frequent and the verification by the researcher is necessary. Besides the automatic tagging, users are also encouraged to realize their manual classification of the expressions according to the taxonomy that best fits their research interests (in the example we distinguished among persons, animal and things).

Finally, the fifth step in an ANTA project is the export of a bipartite graph of expressions and documents. In the resulting network, the link between an expression and the document in which it appears is weighted according to the value of the frequency value for that expression in that document (see figure 6).

FIG. 6. This network shows all the eight Grimm stories analyzed and their most relevant expressions. The size of the nodes entities is proportional to the number of documents in which the expression appears.

The network exported by ANTA can be opened in any network analysis software (i.e. Gephi.org) and "spatialized" with any force-vector algorithm [in the example we used ForceAtlas2, see Jacomy, Heymann, and Venturini 2011]. All tags the researcher has attached to both entities and documents become node attributes and can be used, for example, to give nodes a different color according to their type.

To be sure, it is for the sake of clarity that we presented the work with ANTA as a linear sequence of operations. In practice, things are obviously more complex. The expressions-documents graph, in particular, should not be conceived as the conclusion of the analysis. Rather, it is an intermediary step of exploration helping researchers to go back to ANTA and refine their filtering and classification.

Conclusions

In this paper we described ANTA, a tool developed at the Sciences Po médialab and meant to provide an operationalization of actor-network theory in the domain of text analysis. As readers have certainly remarked, this is still work in progress. Expression extractions should be improved and implemented on open source software. The careful use of natural language processing algorithms could provide better filtering metrics and support in expression merging. Documentations, tutorials and use examples should be supplied. Performances could be improved and integration with other software for text manipulation and analysis should be provided.

More than anything else, the quali-quantitative nature of ANTA needs to be reinforced. Favoring manual investigation over automatic computation and network analysis over statistical aggregations, we managed to remain close to raw data, while scaling to corpora of thousands of texts. Much work remains to be done, however, before ANTA could truly provide a seamless navigation from the global graph of occurrences to the single occurrence of an expression in a specific context. Until then, ANTA remains a handy tool for text analysis, a tool oriented to medium size corpora preferably composed by relatively short documents and particularly suitable for qualitative researchers interested in network analysis.

References

Bastian, M., Heymann, S., and Jacomy, M.,

2010 "Gephi: An Open Source Software for Exploring and Manipulating Networks." In *Proceedings of the Third International ICWSM Conference*.

Bijker, W., and Law, J. (Eds.).

1992 *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: MIT Press.

Callon, M., Law, J., and Rip, A.

1986 *Mapping the Dynamics of Science and Technology*. London: Macmillan.

Callon, M., Millo, Y., and Muniesa, F. (Eds.).

2007 *Market Devices*. Wiley-Blackwell.

Garfinkel, H.

1967 *Studies in Ethnomethodology*. Englewood Cliffs: Prentice Hall.

Glaser, B.G., and Strauss, A.L.

1965 *Awareness Of Dying*. New Brunswick, NJ: Transaction Publishers.

1967 *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New Brunswick, NJ: Transaction Publishers.

Greimas, A.J., and Courtés, J.

1979 *Sémiotique: Dictionnaire raisonné de la théorie du langage, vol. 1*. Paris: Hachette.

Jacomy, M., Heymann, S., and Venturini T.

2011 *ForceAtlas2, A Graph Layout Algorithm for Handy Network Visualization*. White paper available at www.medialab.sciences-po.fr/publications/Jacomy_Heymann_Venturini-Force_Atlas2.pdf

Latour, B.

1998 "Thought Experiments in Social Science: from the Social Contract to Virtual Society." 1st *Virtual Society?* Annual Public Lecture. Brunel University, London, April 1, 1998.

2002 *La fabrique du droit: Une ethnographie du Conseil d'Etat*. Paris: La Découverte.

2005 *Reassembling the Social*. Oxford: Oxford University Press.

Latour, B., and Weibel, P.

2005 *Making Things Public: Atmospheres of Democracy*. Cambridge, MA: MIT Press.

Latour, B., and Woolgar, S.

1979 *Laboratory Life. The Construction of Scientific Facts*. Los Angeles: Sage.

Law, J., and Hassard, J. (eds.)

1999 "Actor Network Theory and After." *Contemporary Sociology* 30: 96.

Marres, N., and Rogers, R.

2005 "Recipe for Tracing the Fate of Issues and their Publics on the Web." In *Making Things Public*, edited by B. Latour and P. Weibel. Karlsruhe/Cambridge, MA: ZKM/MIT Press.

Mol, A.

2001 *The Body Multiple: Artherosclerosis in Practice*. Durham and London: Duke University Press.

Muniesa, F., and Callon, M.

2007 "An Introduction to Market Devices." In *Market Devices*, edited by M. Callon, Y. Millo, and F. Muniesa. Oxford: Blackwell Publishers.

Rogers, R.

2005 *Information Politic on the Web*. Cambridge, MA: MIT Press.

2009 *The End of the Virtual: Digital Methods*. Amsterdam: Amsterdam University Press.

Rogers, R., and Marres, N.

2002 "French scandals on the Web, and on the streets: A small experiment in stretching the limits of reported reality." *Asian Journal of Social Science* 66: 339-353.

Teil, G., and Latour, B.

1995 "The Hume Machine: Can Association Networks Do More than Formal Rules?" *Stanford Humanities Review* 4:47-66.

Tukey, J.W.

1977 *Exploratory Data Analysis*. Reading: Addison-Wesley.

Venturini, T.

2012 "Great expectations: méthodes quali-quantitative et analyse des réseaux." In *L'Ère post-média. Humanités digitales et Cultures numériques*, edited by J.-P. Fourmentraux. Paris: Hermann.

Venturini, T., and Latour, B.

2010 "The Social Fabric: Digital Traces and Quali-quantitative Methods." In *Proceedings of Future en Seine*. Paris: Editions Futur en Seine.

Woolgar, S. (ed.)

2002 *Virtual Society?: Technology, Cyberbole, Reality*. Oxford: Oxford University Press.

Once Upon a Text: an ANT Tale in Text Analysis

Abstract: ANTA or Actor-Network Analyzer is a simple piece of software developed at Sciences Po médialab to offer social researchers a simple text-analysis tool attuned with the theoretical tenets of actor-network theory. Striving to make actor-network theory compatible with modern text-analysis, we have learned much about both. In this paper we'll discuss our adventure in ANT and text-analysis while describing the basic functions of ANTA and providing examples of its usage.

Keywords: Actor-network theory, text analysis, digital methods, network analysis

Tommaso Venturini is associate professor at Sciences Po Paris where he coordinates the research activities of the médialab and is responsible for the courses of Controversy Mapping.

Daniele Guido works as interface and information designer at Sciences Po médialab and is the principal developer of ANTA