Zeitschrift: Horizons : le magazine suisse de la recherche scientifique

Herausgeber: Fonds National Suisse de la Recherche Scientifique

Band: 27 (2015)

Heft: 105

Rubrik: Point fort humanités numériques : des chiffres et des lettres

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 24.08.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Des chiffres et des lettres

Les sciences humaines adoptent elles aussi les technologies numériques. Ces outils offrent un nouveau regard sur notre culture et obligent les chercheurs à redéfinir leur discipline.

Le numérique entre dans l'histoire

Historiens et archéologues apprennent à apprivoiser les outils informatiques. Une évolution inéluctable qui a le potentiel de bouleverser ces disciplines.
Par Fabien Goubet

Le monde se numérise et les sciences humaines n'y échappent pas. Elles aussi se métamorphosent sous l'influence des ordinateurs et des algorithmes. Il ne s'agit évidemment pas de simples outils de traitement de texte ou de messagerie, mais bien de profondes mutations qui s'immiscent dans toutes les étapes de production des savoirs, de la numérisation des sources à leur analyse. Emergence de nouvelles pratiques, apparition de nouveaux métiers, adaptation nécessaire des formations universitaires: les humanités numériques ont tout pour rebattre les cartes des sciences humaines.

Le terme «digital humanities» n'est apparu officiellement à l'écrit qu'en 2004, précise Claire Clivaz, professeure invitée au Laboratoire de culture et humanités digitales de l'Université de Lausanne: «On parlait auparavant de ‹humanities and computing› comme si ces deux notions se faisaient face.» Le changement sémantique montre qu'elles sont aujourd'hui intimement liées.

L'ordinateur dessine les frontières

Les travaux de Sylvian Fachard, de l'Université de Genève, illustrent bien les nouvelles possibilités offertes par le numérique. Ce spécialiste de l'archéogéographie veut déterminer le tracé des frontières des différentes communes de l'Attique, la région entourant Athènes, entre le Xe et le Ier siècle av. J.-C. «Les quelques rares textes qui nous sont parvenus ont tous été analysés au début du XXe siècle et ne permettent pas de replacer les frontières sur une carte», explique Sylvian Fachard. Face à cette impasse, le chercheur recourt à la modélisation spatiale de la région. Il répertorie en fonction de l'espace et du temps tous les objets, ruines et autres tombeaux découverts lors de fouilles, et croise ces informations avec des données géographiques telles que la topographie du terrain. En fonction de ces données, un algorithme estime avec une certaine probabilité l'emplacement des frontières à telle ou telle époque.

Ces outils numériques ne dispensent pas le chercheur de son travail de terrain, souligne Sylvian Fachard. Une fois les prédictions livrées par le modèle, il remet son chapeau d'archéologue et cherche des indices pour les confirmer ou les infirmer. «Il est nécessaire d'interpréter toutes les données afin de produire un résultat pertinent. Le numérique génère encore davantage d'informations qu'il faut articuler entre elles. Cela réclame un niveau d'analyse plus complexe.»

L'histoire collective

Autre profonde mutation induite par le numérique, le changement d'échelle. «Dans la culture digitale, on passe du solitaire au collectif, comme l'ont fait les sciences dures, note Claire Clivaz. Les articles sont signés par des collaborations de chercheurs et non plus par des auteurs uniques.»

La composante temporelle est également affectée, ajoute Sylvian Fachard: «Nous partons en fouilles équipés de tablettes tactiles qui nous permettent de centraliser en temps réel les données récoltées. Nous pouvons adapter notre stratégie au jour le jour. Cela nous apporte une souplesse incroyable.» «Le rythme de la science évolue, ajoute Claire Clivaz. La production est assurée en continu, notamment avec les blogs sur lesquels s'expriment les chercheurs.»

> «Il n'y a plus un texte unique dans une version donnée, mais une véritable généalogie de toutes les versions.»

Claire Clivaz étudie les manuscrits du Nouveau Testament, et le numérique permet de mettre enfin la main sur une partie des quelque 5800 manuscrits rédigés en grec ancien. Mais le numérique change également la manière de penser ce qu'est un document historique. «Il n'y a plus un texte unique dans une version donnée, mais une véritable généalogie de toutes les versions, note la chercheuse. Le texte devient par conséquent multiforme. Cela change complètement notre état d'esprit: avant, on cherchait la version la plus ancienne à tout prix, tandis qu'on perçoit maintenant un texte comme évoluant dans une histoire de la lecture.»

Citant l'historien du livre Roger Chartier, la chercheuse insiste sur le fait que le support de l'information modèle notre façon de penser. Et d'ajouter: «Plus encore

◂ Pages 11 et 12. Représenter la vie de Newton, de Washington et d'autres géants de l'histoire par un portrait ou sur une carte? Une équipe emmenée par Dirk Helbing, d'ETH Zurich, a relié les lieux de naissance et de mort de 120 000 personnalités au cours des 2600 dernières années pour illustrer les déplacements des centres intellectuels.
Images: Wikimedia Commons, Wellcome Library (p. 11); M. Schich et M. Martino (2014) (p.12)

que l'imprimerie, la révolution numérique est la rupture la plus importante depuis le passage du rouleau de parchemin au codex (un cahier composé de pages manuscrites, ndlr)».

Générer des hypothèses

Cette rupture est notamment marquée par un changement de relation entre les historiens et leurs sources. Martin Grandjean, un doctorant en histoire à l'Université de Lausanne, planche sur les archives de la Société des Nations afin de comprendre comment les chercheurs allemands actifs durant l'entre-deux-guerres ont réussi à revenir au premier plan européen malgré leur mise à l'écart. Il élabore pour cela une cartographie spatiotemporelle de leur correspondance.

> «L'un des défis est de savoir jusqu'où collaborer avec des personnes issues d'autres disciplines sans perdre notre identité.»

«Le contenu des lettres m'intéresse moins que les métadonnées, à savoir qui a écrit à qui et à quelle date», explique le chercheur. Il esquisse ensuite des représentations schématiques des métadonnées dans l'espoir de faire émerger des informations qui auraient jusqu'ici échappé aux historiens. Mais aussi riche que puisse être sa cartographie, elle ne représente pas un résultat en tant que tel: «La visualisation de données me donne des idées et me suggère des hypothèses. Comme n'importe quel historien, je dois ensuite les vérifier en me plongeant dans les archives.»

Ces visualisations de données complexes se trouvent également au cœur d'un projet de recherche des universités de Berne et de Giessen, le Repertorium Academicum Germanicum (RAG). Aux racines de celui-ci, une base de données sur 50 000 savants et intellectuels ayant fréquenté les universités du Saint-Empire romain germanique entre 1250 et 1550. En principe, rien n'interdirait d'établir ce «who's who» sans outil numérique, mais il serait bien plus difficile d'interpréter les résultats. Les chercheurs ont choisi de les présenter sur une carte en fonction du temps, une visualisation originale qui, là aussi, peut faire naître des nouvelles hypothèses. «Cet outil nous permet de repenser le rôle des universités médiévales dans une perspective beaucoup plus large, commente Kaspar Gubler, directeur adjoint du RAG. La nature numérique de notre index rend également possible l'établissement de davantage de liens avec d'autres projets de recherche.»

Big data du passé

La puissance des outils numériques peut faire rêver, mais des promesses emphatiques sont aussi susceptibles d'attiser les critiques. Venice Time Machine, un projet médiatique de l'EPFL, ambitionne de recréer la Cité des Doges des douze derniers siècles à partir des archives d'Etat vénitiennes. La première étape consistera à numériser les millions de documents (quelque 80 kilomètres d'archives) existants. «Analyser autant de documents serait tout bonnement impossible pour des êtres humains», avance Frédéric Kaplan, chef d'orchestre du projet et professeur au Laboratoire d'humanités digitales de l'EPFL.

Ce seront donc des machines qui devront déchiffrer les écritures manuscrites et en extraire des données telles que le prix de la cannelle importée d'Orient, le salaire des condottieri ou encore la liste des apprentis maçons à une époque donnée. «En résumé, ce projet, c'est faire du big data du passé pour créer les Google Maps et le Facebook de Venise», lance Frédéric Kaplan. Il compare son projet au séquençage du génome humain, «une grande infrastructure qui permettra d'aider de nombreux chercheurs par la suite». Avec ses partenaires vénitiens, le chercheur espère même mettre sur pied une visite touristique virtuelle de la Sérénissime sur smartphone.

Cette vision grandiose attise le scepticisme, voire la méfiance, de certains historiens qui préfèrent ne pas être cités. Leur critique: le projet manquerait de cadrage et nécessiterait des approches historiques complémentaires. Mais Frédéric Kaplan reste sûr de lui: «Nous publierons bientôt une modélisation complète d'un premier quartier de Venise, celui du Rialto.»

Garder son identité

Les humanités numériques resteront-elles de simples outils dans la trousse des historiens, ou annoncent-elles l'avènement d'une discipline à part entière? «C'est ce que je pense, mais cette question est sujette à débat au sein de la communauté», répond Frédéric Kaplan, alors que Sylvian Fachard réaffirme «rester un archéologue avant tout». Claire Clivaz nuance: «Pour moi, ce n'est ni l'un ni l'autre. Je vois plutôt les humanités numériques comme un label de transition, car en définitive les sciences humaines et sociales seront numériques quoi qu'il arrive. Si bien que les adjectifs ‹numérique› et ‹digital› pourraient bien disparaître d'eux-mêmes, par évidence.»

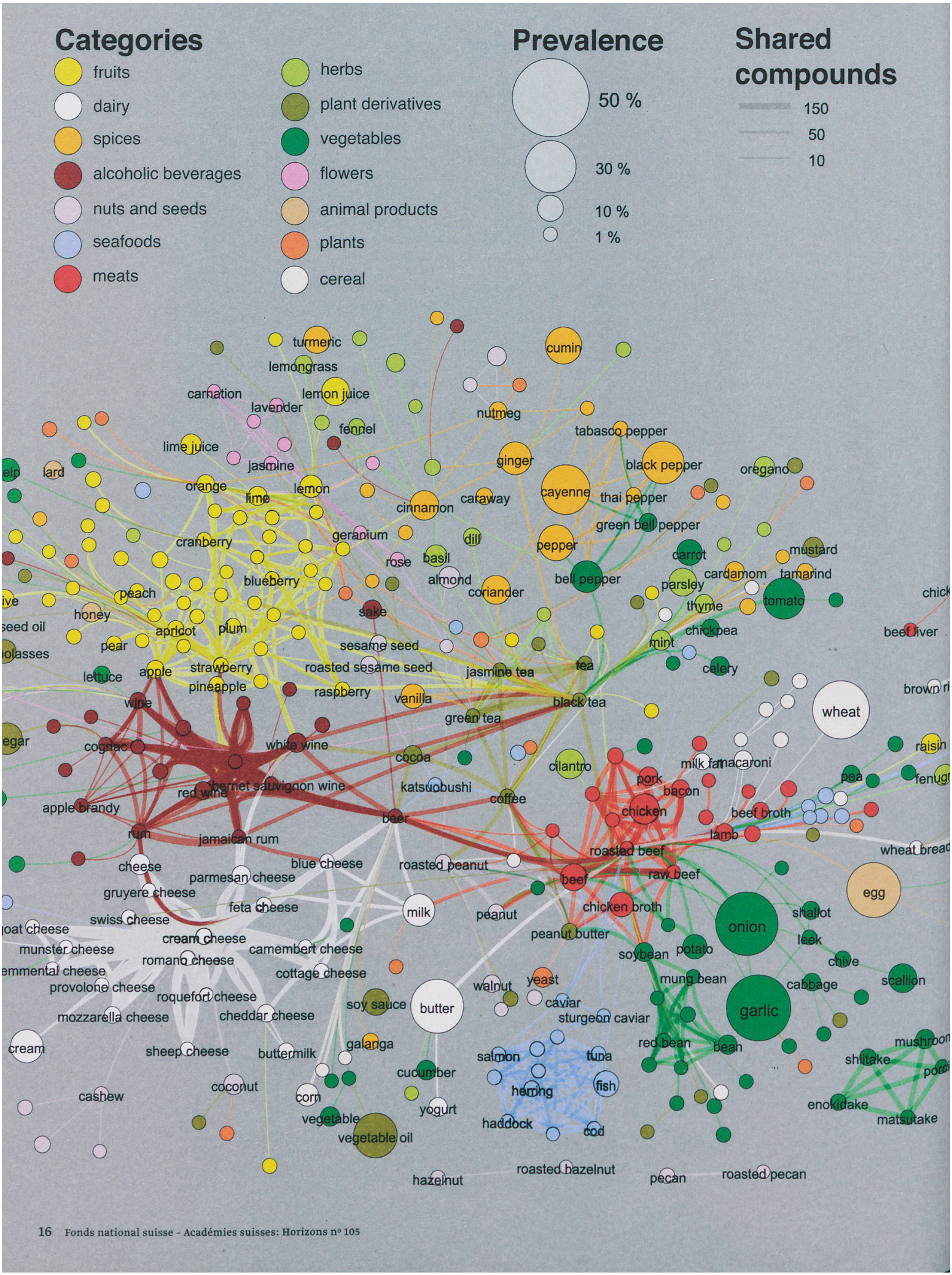
Les passerelles que lance le numérique ne sont pas toujours faciles à fréquenter. Dans les couloirs de ces nouveaux centres se croisent désormais historiens, sociologues, informaticiens, statisticiens. «Nous avons besoin de profils capables de faire le lien entre tous ces univers», souligne Bella Kapossy, du Laboratoire de cultures et humanités digitales de l'Université de Lausanne. Toute évolution d'une discipline amène une redéfinition potentielle de son identité. Comme le dit Claire Clivaz: «L'un des défis posés par cette révolution est de savoir jusqu'où collaborer avec des personnes issues d'autres disciplines sans perdre notre identité.»

Fabien Goubet est journaliste scientifique et écrit pour Le Temps.

Les universités suisses en mutation

Onze ans après son apparition, l'expression «digital humanities» commence à se faire une place dans le cursus académique. En Suisse, étudiants et chercheurs ont à leur disposition des cours, des ateliers et des séminaires ponctuels. Depuis cette année, la Faculté des lettres de l'Université de Lausanne (UNIL) propose également 30 crédits de spécialisation en digital humanities au niveau master.
Le réseau européen Dariah répertorie tous les cours disponibles, sur une carte en ligne, bien évidemment. Une douzaine d'entrées suisses y figurent, réparties entre les universités de Genève, Lausanne, Berne et Bâle, partenaires de cette infrastructure. Le paysage helvétique va encore évoluer dans les mois à venir, notamment avec un projet de Master interfacultaire en humanités digitales à l'UNIL.

PATAK'S
Meerrettich
KARLSBRAU
500 g
Mozzarell


Categories
fruits
dairy
spices
alcoholic beverages
nuts and seeds
seafoods
meats
herbs
plant derivatives
vegetables
flowers
animal products
plants
cereal
Prevalence
50 %
30 %
10 %
1 %
Shared compounds
150
50
10
turmeric
lemongrass
cumin
carnation
lemon juice
lavender
nutmeg
fennel
tabasco pepper
lime juice
jasmine
ginger
black pepper
oregano
orange
lime
lemon
cinnamon
caraway
cayenne
thai pepper
green bell pepper
cranberry
geranium
dill
pepper
carrot
mustard
rose
basil
almond
bell pepper
cardamom
tamarind
blueberry
coriander
parsley
peach
sake
thyme
tomato
honey
apricot
plum
chickpea
pear
sesame seed
mint
beef liver
lettuce
apple
strawberry
roasted sesame seed
jasmine tea
tea
celery
pineapple
raspberry
vanilla
wine
black tea
wheat
green tea
cognac
white wine
cocoa
cilantro
milk
macaroni
raisin
pea
pork
bacon
cabernet sauvignon wine
katsuobushi
red wine
coffee
chicken
beef broth
apple brandy
beer
lamb
rum
jamaican rum
roasted beef
wheat bread
cheese
blue cheese
roasted peanut
parmesan cheese
beef
raw beef
egg
gruyere cheese
feta cheese
milk
peanut
chicken broth
onion
shallot
swiss cheese
cream cheese
peanut butter
leek
munster cheese
camembert cheese
soybean
potato
romano cheese
chive
cottage cheese
mung bean
cabbage
scallion
provolone cheese
roquefort cheese
walnut
yeast
soy sauce
butter
caviar
garlic
mozzarella cheese
cheddar cheese
sturgeon caviar
cream
sheep cheese
buttermilk
galanga
red bean
bean
mushroom
salmon
tuna
shiitake
cucumber
coconut
fish
cashew
corn
herring
enokidake
yogurt
vegetable
haddock
cod
matsutake
vegetable oil
roasted hazelnut
hazelnut
pecan
roasted pecan

200 ans de littérature en 0,4 seconde

La lecture lente a vécu: des ordinateurs avalent des millions de livres en un rien de temps. Et nous proposent leur propre analyse. *Par Mirko Bischofberger*

Tout a commencé avec des données. Beaucoup trop nombreuses et trop complexes. En 1940, le jésuite Roberto Busa s'attelle à produire un index complet des 11 millions de mots retrouvés dans les écrits théologiques de Thomas d'Aquin. Une entreprise gigantesque, pour laquelle une vie entière ne suffirait pas. Mais le père Busa a une idée: se faire aider par une machine. Il trouve le soutien de Thomas Watson, le fondateur d'IBM, et après trois décennies vient à bout de son projet: 56 volumes et 70 000 pages. L'Index Thomisticus est la première œuvre à permettre une recherche simple et rapide dans les contenus d'un corpus entier.

La montagne et le sport

La numérisation s'immisce aujourd'hui dans toutes les sciences humaines. «La linguistique et la littérature sont les principales intéressées, explique Martin Volk, professeur de linguistique computationnelle à l'Université de Zurich. Un accès numérique à leurs données de recherche leur permet d'étayer ou d'invalider certaines hypothèses à l'aide de chiffres et de statistiques.» Dans le cadre de son propre projet de recherche Text+Berg, le spécialiste a numérisé les 250 volumes du Club Alpin Suisse (CAS) parus depuis 1864. «Ce matériel est une mine d'informations. Il montre comment la façon d'envisager les montagnes a évolué avec le temps. Décrites autrefois comme des objets d'exploration, elles sont vues aujourd'hui comme un terrain d'entraînement sportif. Le terme ‹compétition› est, par exemple, beaucoup plus fréquent que par le passé.»

A l'Université de Genève, des chercheurs veulent numériser une partie de la Bibliotheca Bodmeriana, une collection exceptionnelle de 150 000 œuvres littéraires recouvrant trois millénaires en 80 langues. On y trouve, entre autres, le plus ancien manuscrit de l'Evangile de Jean datant du IIe siècle ainsi que les originaux des contes des frères Grimm.

Le citoyen à l'aide

Mais numériser des ouvrages est une tâche pénible. «Il faut couper les livres à la main avant de scanner chaque page séparément, détaille Martin Volk, qui a numérisé dans son projet plus de 120 000 pages. On a alors uniquement des images, mais pas de texte.» La reconnaissance des textes est effectuée par des programmes qui identifient les lettres dans les images et les transforment en mots. «Mais le taux d'erreurs reste encore assez élevé, notamment avec des écrits anciens du XIXe siècle.». Dans son projet, ce taux était de douze erreurs par page. Il a tout fallu vérifier à la main.

«L'outil Ngram joue un rôle pionnier pour les sciences humaines.»

Martin Volk a alors développé un système de correction en ligne qui a permis à des volontaires d'éliminer les erreurs, sous forme de jeu. Ce projet de science citoyenne a séduit les membres du CAS. «Grâce à leur aide, nous avons pu effectuer plus de 250 000 corrections en six mois.» Le corpus numérique est aujourd'hui presque correct à 100%. Une fois numérisés, les textes peuvent être facilement archivés et consultés, une chose «impossible pour des documents anciens, rares ou difficilement accessibles», souligne le chercheur.

Freud, Einstein, Darwin

Google Books est sans doute l'archive de ce genre la plus célèbre et la plus complète. Sa recherche en texte intégral permet de parcourir les stocks des bibliothèques universitaires de Harvard, Stanford et New York. Certaines bibliothèques européennes sont également intégrées, comme celle de l'Université d'Oxford ou la Bayerische Staatsbibliothek.

Ce gigantesque corpus a donné naissance en 2010 à Google Ngram, une application web qui analyse la fréquence d'un mot ou d'une suite de mots dans tous les

◂ Pages 15 et 16. Les deux nous parlent de ce que nous mangeons: un regard jeté dans notre frigo et une visualisation analytique des arômes. La couleur indique les catégories, la taille des nœuds représente leur fréquence dans les recettes, et les liens le nombre de composants aromatiques qu'ils partagent.
Images : Valérie Chételat (p. 15); Yong-Yeol Ahn (p. 16)

ouvrages dès 1800 scannés par Google. Elle permet d'étudier des événements historiques, comme l'abolition de l'esclavage, mais aussi d'observer l'évolution linguistique de certains mots au sein d'une langue, ou encore la popularité des personnalités: les célébrités scientifiques que sont Freud, Einstein ou Darwin apparaissent toutes très fréquemment dans la littérature, mais Freud est cité deux fois plus souvent depuis 1950.

«Ngram n'est qu'un exemple de ce qu'on peut faire aujourd'hui avec les données culturelles numérisées», remarque son créateur Jean-Baptiste Michel, chercheur à l'Université Harvard. Les sciences humaines numériques sont aujourd'hui inconcevables sans Ngram, dont Martin Volk confirme le «rôle pionnier».

La culture du SMS

La numérisation de la littérature n'est qu'une approche d'analyse linguistique. «Avec nos ordinateurs et téléphones, nous saisissons plus de textes numériques que jamais», souligne Elisabeth Stark, du Séminaire de langues romanes de l'Université de Zurich. Rien qu'en 2013, plus de 100 millions de SMS ont été envoyés chaque jour en Allemagne. «Ces textes ne sont presque jamais imprimés, mais font partie de notre culture linguistique», poursuit la chercheuse. Avec son projet du Fonds national suisse Sms4science, elle étudie les caractéristiques linguistiques de la communication par SMS en Suisse.

La langue raccourcie des SMS obéit aux même lois que celles du langage parlé.

Pour accéder à ces données, Elisabeth Stark et ses collègues ont invité les utilisateurs de portables en Suisse à envoyer une copie de leurs SMS à un numéro gratuit. «Nous avons ainsi pu collecter environ 26 000 messages.» Son équipe s'intéresse notamment aux ellipses linguistiques, autrement dit aux omissions de mots, comme dans les expressions «Arrive bientôt» ou «T'appelle». Afin de découvrir pourquoi le sujet est omis dans ces exemples, les chercheurs ont analysé tous les SMS en français et en allemand. Résultat: ces omissions sont beaucoup plus rares qu'on ne l'imaginait et obéissent aux mêmes lois que le langage parlé quotidien. «Cela contredit l'impression que l'on a souvent lorsqu'on regarde des SMS isolés, explique Elisabeth Stark. D'où la nécessité de disposer d'une grande quantité de données.»

Accéder aux données

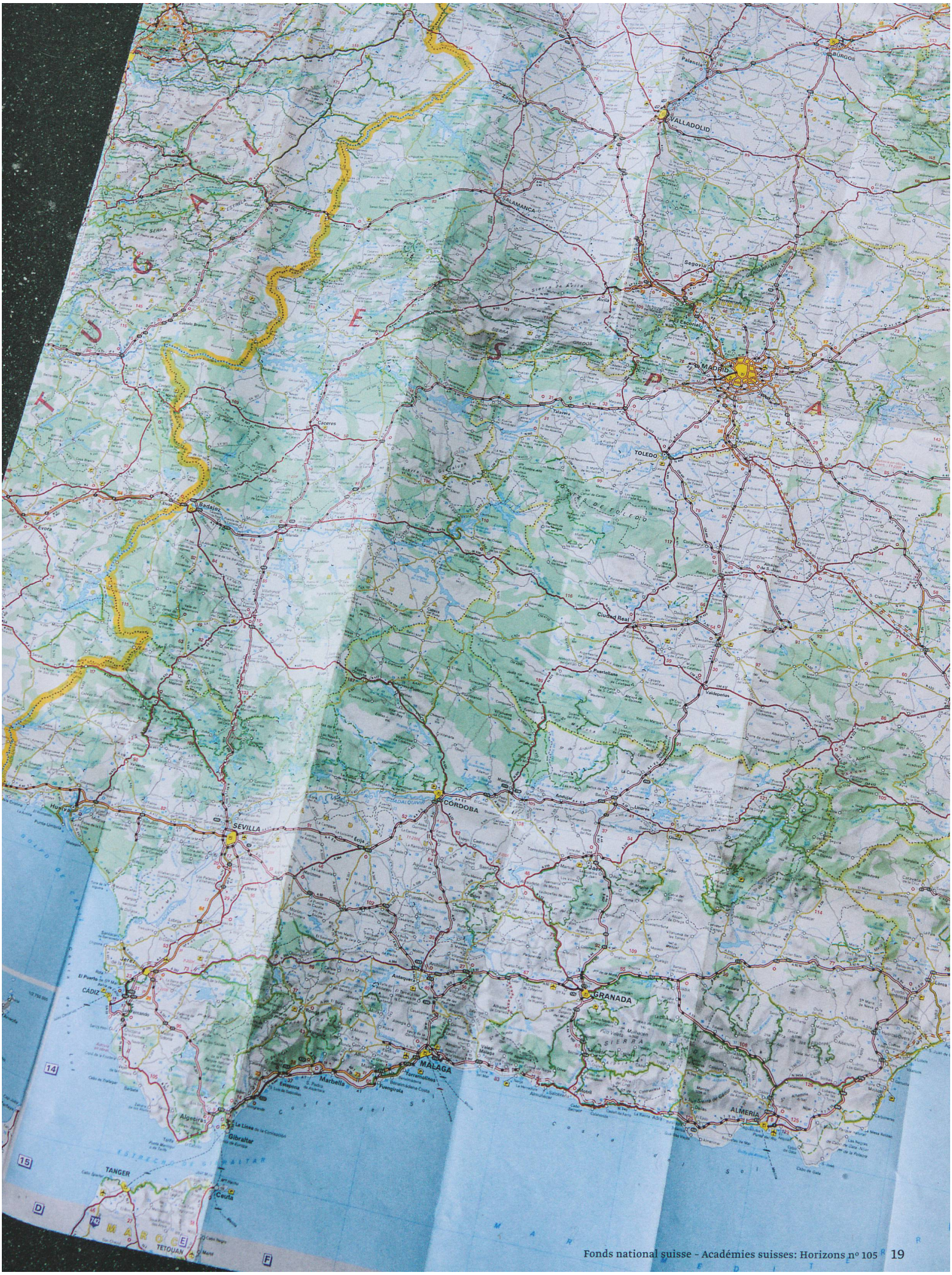
Les sciences humaines numériques permettent d'analyser la littérature et la langue à l'aide de chiffres, qui ont toujours été la marque des sciences exactes. Ils décrivent les schémas quantitatifs et les relations avec une précision dont les mots sont rarement capables. La prochaine génération de chercheurs en sciences humaines travaillera sur des données, comme le font les bioinformaticiens depuis la fin du XXe siècle. «Ce domaine profitera de l'augmentation massive de la quantité de textes numérisés, opine Martin Volk. Comme le séquençage du génome, qui a conduit à la bioinformatique, la numérisation de notre langue et de notre littérature fera inévitablement bientôt partie intégrante des sciences humaines.»

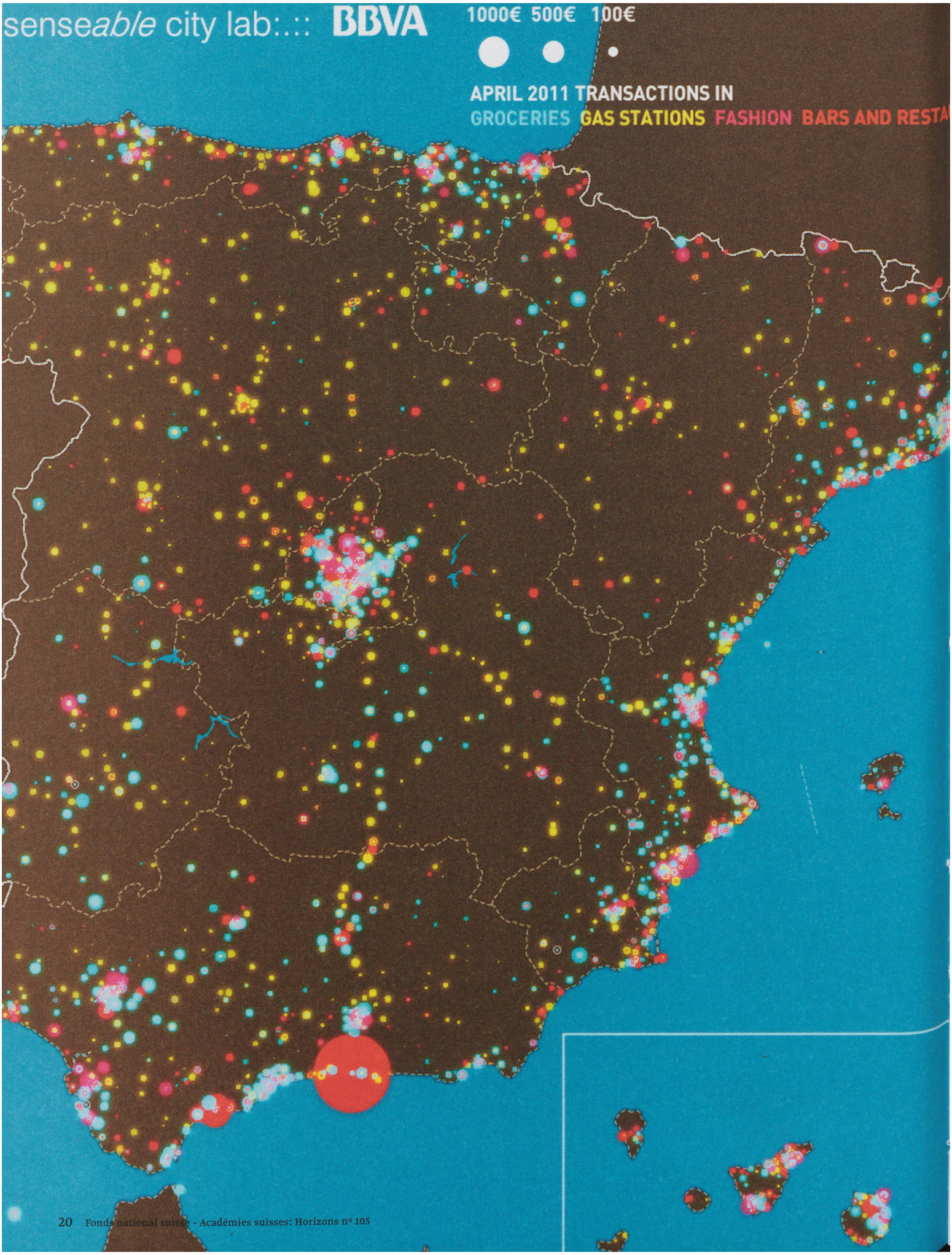
Des chercheurs comme Martin Volk et Elisabeth Stark ne représentent que le début d'une nouvelle ère dans la recherche. «Malheureusement, les ressources allouées aux sciences humaines numériques sont limitées pour l'instant, en Suisse», déplore Martin Volk. «Dans toute l'Université de Zurich, par exemple, il n'y a pas encore de chaire d'humanités numériques alors qu'il serait grand temps», renchérit Elisabeth Stark. Mais les deux chercheurs semblent trouver encore plus important de pouvoir accéder aux principaux réservoirs de données. «Des initiatives européennes d'envergure existent, mais la Suisse n'en fait malheureusement pas partie pour l'instant», regrette Elisabeth Stark. Jean-Baptiste Michel abonde dans son sens: «Pouvoir accéder aux données constitue le moteur essentiel!»

Mirko Bischofberger est collaborateur scientifique du FNS.

—

J.-B. Michel et al., Quantitative Analysis of Culture Using Millions of Digitized Books. Science, 2011


PORTUGAL
ESPAÑA
BURGOS
Palencia
VALLADOLID
SALAMANCA
Segovia
Avila
El Escorial
MADRID
Guadalajara
Cáceres
TOLEDO
Badajoz
Ciudad Real
Valdepeñas
CORDOBA
SEVILLA
Huelva
Jaen
Antequera
GRANADA
MALAGA
Marbella
Estepona
Torremolinos
Fuengirola
ALMERIA
Jerez
El Puerto
CADIZ
Algeciras
Gibraltar
TANGER
Ceuta
TETOUAN
MARRUECOS
14
15
D
F


senseable city lab:.:: BBVA
1000€ 500€ 100€
APRIL 2011 TRANSACTIONS IN
GROCERIES GAS STATIONS FASHION BARS AND RESTA

Les archéologues du numérique

Numériser les objets historiques ne les préserve pas du déclin. Un projet suisse veut sauver les données produites par la recherche.
Par Daniel Saraga

Internet n'oublie jamais», assènent les experts inquiets de voir notre vie privée étalée sur le web. Mais rien n'est moins sûr: le changement constant des supports de l'information numérique et des formats des fichiers laisse derrière lui des montagnes d'informations qui, inexorablement, deviennent illisibles. Livrées à elles-mêmes, elles sont condamnées à l'oubli.

Cette problématique touche de plein fouet les résultats scientifiques, exprimés de plus en plus souvent sous forme de bases de données. Car celles-ci vieillissent très rapidement: les langages de programmation se révèlent obsolètes et les systèmes d'exploitation ne sont plus compatibles avec le nouveau serveur.

«C'est un vrai problème, souligne Lukas Rosenthaler, responsable du Centre de données et services pour les sciences humaines (DDZ), un projet consacré à assurer la pérennité des résultats scientifiques en sciences humaines (voir «Sauvegarder le sauveur»). Il est très rare que les chercheurs continuent à maintenir les infrastructures numériques une fois un projet - et son financement - terminé. Une base de données inaccessible est inutilisable, et ne pas l'entretenir revient à détruire une production scientifique. Paradoxalement, ce type de recherche numérique peut s'avérer bien plus fragile qu'un résultat publié sous forme d'article.»

Les copistes du IIIe millénaire

Le directeur des Digital Humanities Lab à l'Université de Bâle a pu sauver l'une des plus importantes bases de données consacrée à la mythologie grecque, le Lexicon Iconographicum Mythologiae Classicae, qui s'est arrêté en 2009, après trente ans de développement: «Elle était complètement hors service, et l'entreprise qui l'avait programmée était en faillite, poursuit Lukas Rosenthaler. Nous avons même dû pirater le site, car les mots de passe avaient disparu. Notre travail ressemble parfois à une sorte d'archéologie du numérique.» Ces efforts redonnent une seconde vie aux résultats de la recherche. L'Université Harvard, par exemple, s'intéresse aujourd'hui à intégrer le Lexicon dans son commentaire d'Homère, en utilisant les «linked open data», une composante du web 3.0 qui permet de relier les informations en ligne de manière directe et dynamique.

Avec sa petite équipe, l'ancien physicien s'est appuyé sur la technologie sémantique pour créer une plateforme générique capable de structurer des données provenant de plateformes très variées. «J'estime que nous pouvons traduire 99% des bases de données utilisées en sciences humaines, et même certains projets en biologie. En trois ans, nous avons migré une trentaine de projets, de la mythologie grecque à une collection de photographies historiques de montagne.»

Le DDZ suit le concept de l'Open Archival Information System: copier les données régulièrement et les retranscrire dans un nouveau format plus actuel. Un processus difficile et coûteux, qu'il faut répéter régulièrement - le pendant digital du travail des moines copistes du Moyen Âge. «La plupart des groupes de recherche n'ont pas les moyens de créer des outils stables, poursuit Lukas Rosenthaler. L'idéal est de travailler avec eux dès le début pour élaborer une base de données durable qui puisse ensuite être facilement mise à jour et migrée.»

Le mouvement de l'Open Research Data encourage les chercheurs à générer des résultats scientifiques accessibles à chacun et de manière interactive. Mais comme toute machine, l'outil exige un entretien régulier. Un travail de l'ombre essentiel pour éviter de voir cette connaissance disparaître à jamais.

Daniel Saraga est responsable de la communication scientifique du FNS.

—

Final report for the pilot project «Data and Service Center for the Humanities» SAGW, 2015
Le patrimoine culturel à l'ère numérique, Nike Bulletin, juin 2014
http://openresearchdata.ch/

Sauvegarder le sauveur

Voué à la préservation des archives numériques, le Centre de données et services pour les sciences humaines (DDZ) est lui-même en danger. «Nous nous battons depuis 2008 pour mettre en place une plateforme stable, déclare Markus Zürcher, secrétaire général de l'Académie suisse des sciences humaines, qui a lancé le projet. Toutes les personnes concernées soutiennent cette plateforme. La seule chose à régler, c'est le financement.»
Le DDZ reste un projet pilote qui touche à son terme. «Nous avons déposé en mars 2015 une demande portant sur 2 millions de francs pour 2017–2020 auprès du SEFRI (Secrétariat d'Etat à la formation, à la recherche et à l'innovation). D'ici là, nous sommes prêts à assurer le financement jusqu'en 2017, car une interruption du projet serait très néfaste.» Par comparaison, quelque 30 millions sont dépensés annuellement pour les bases de données dans les sciences humaines.

◂ Pages 19 et 20. La cartographie créé une forme de réalité augmentée en ajoutant des données à la géographie. Ces vues globales permettent d'identifier rapidement des informations cachées, comme ici l'explosion d'achats alimentaires avant Pâques et celle de la restauration sur la Costa Tropical au sud de l'Espagne.
Images: Valérie Chételat (p. 19); MIT Senseable City Lab – Carlo Ratti, Director – Pedro Cruz: Visualization (p. 20)

«Il ne s'agit pas d'une révolution»

Les humanités numériques ne constituent que des nouvelles méthodes, mais elles permettent de poser un nouveau type de questions. Gerhard Lauer est un pionnier des techniques informatiques pour l'analyse littéraire qui se décrit comme un conservateur.
Propos recueillis par Urs Hafner

Notre héritage culturel ne devrait pas être défini par Google, soutient Gerhard Lauer.

Gerhard Lauer, que comprenez-vous par le terme d'humanités numériques?
Quelque chose de très pratique: l'utilisation de méthodes informatiques pour réaliser des éditions numériques et analyser de gros volumes de textes de manière plutôt quantitative.

Les humanités numériques ont-elles modifié votre recherche?
Oui, mais pas abruptement. Elles commencent seulement à transformer la recherche en sciences humaines, dans la mesure où elles étendent peu à peu le spectre des méthodes et des questions. Nous entreprenons l'analyse de la littérature de manière quantitative, par exemple celle du «Werther» de Goethe ou les récits de Kafka. Du côté des étudiants, la demande pour ces méthodes augmente, et j'accompagne à présent les premiers travaux de bachelor dans ce domaine.

Traditionnellement, le texte se trouve au cœur des sciences humaines. Celles-ci analysent son sens manifeste et latent, et ouvrent leurs conclusions à la discussion sur un mode narratif et argumentatif. Les humanités numériques passent à côté de cette dimension.
A première vue, vous avez raison: ce qu'elles font avant tout, c'est compter des mots. Mais la répartition de ces derniers dans les textes dit bel et bien quelque chose sur ces mêmes textes. Le type de mots utilisé par un auteur nous en apprend beaucoup sur lui. On peut penser que l'emploi d'articles et de pronoms ne dit pas grand-chose, mais c'est le contraire.

Quelles connaissances avez-vous pu ainsi générer?
Nous nous intéressons par exemple au nombre de mots courts et de mots longs qu'utilisait Kafka, comparé à d'autres auteurs de son époque, ou encore à son usage de certains mots fonctionnels. La répartition statistique des fréquences de mots dans ses récits permet d'objectiver la particularité de son style. Elle nous dit aussi quelque chose sur l'époque. Nous n'étudions plus seulement le canon littéraire, comme «Les Affinités électives» de Goethe, mais les nombreux livres qui étaient populaires à la même période. L'histoire culturelle de ce qui était lu passe ainsi au premier plan, et plus seulement les belles lettres. Il s'agit de premiers éléments, pas encore de connaissances qui chamboulent tout. La nouveauté, ce sont les méthodes.

L'analyse historico-herméneutique n'est-elle pas supérieure au comptage de mots?
Tel est peut-être encore le cas, mais cela change au fur et à mesure des résultats que nous obtenons. Nous identifions toujours plus clairement certains schémas textuels à partir desquels nous pouvons déduire l'évolution de l'aptitude humaine à raconter. On peut ainsi aborder de nouvelles questions, notamment la différence entre narration européenne et narration asiatique.

Comment procédez-vous lorsque vous cherchez à savoir ce que les gens lisaient en Europe au début du XIXe siècle?
Nous évaluons des sources et des métadonnées recueillies à partir de différents catalogues de bibliothèques. De là, nous pouvons déduire quels livres étaient imprimés, achetés, empruntés et probablement lus par les gens. Ou alors, nous utilisons l'outil Ngram Viewer de Google qui permet de parcourir près de cinq millions de livres en différentes langues.

L'analyse Ngram est pourtant sélective: les critères en fonction desquels Google numérise ou non certains livres restent inconnus.
C'est vrai. Les corpus avec lesquels nous travaillons n'ont souvent pas été constitués de manière systématique et ne sont pas équilibrés statistiquement. Google Books est particulièrement problématique, car Google a numérisé au hasard des bibliothèques entières. Ce n'est pas un corpus. Cette entreprise définit ainsi ce qui constitue notre héritage culturel. Il incombe donc aux universités et aux bibliothèques de s'engager pour que l'héritage culturel ne soit pas privatisé, mais qu'il puisse être évalué de façon critique. En sciences humaines, nous faisons trop peu face à ce défi culturel.

Les adeptes des humanités numériques parlent souvent de révolution. Vous aussi?

C'est la numérisation qui est une révolution, pas les humanités numériques. Les sciences humaines se transforment, comme la chimie, la physique, la médecine et la biologie se sont transformées lorsqu'elles ont intégré certaines méthodes informatiques. La biologie a profondément changé avec le recours à l'ordinateur, mais sans rien perdre en tant que discipline. Il se produira la même chose avec les sciences humaines. L'archéologie et les sciences du langage ont déjà franchi le pas. Les humanités numériques représentent pour ainsi dire un sentiment de révolution: quelque chose de nouveau arrive, notamment des chiffres et des statistiques, et de nombreuses disciplines ne savent pas encore comment l'appréhender.

«Nous ne disposons pas encore de méthodologie confirmée.»

Les sciences humaines prennent-elles pour modèle les sciences naturelles, qui sont habituées aux méthodes quantitatives, dans l'espoir d'obtenir davantage de fonds de recherche?

Il faut distinguer entre deux intérêts contradictoires. D'un côté, la politique d'encouragement de la recherche s'oriente en fonction des tendances vers ce qui promet la prochaine nouveauté. Actuellement, les humanités numériques sont considérées comme un domaine prometteur, on y investit donc de l'argent. D'un autre côté, il s'agit d'un secteur où il est difficile de trouver un emploi dans les facultés établies. Celles-ci font souvent preuve de beaucoup de retenue et ont plutôt tendance à engager un chercheur qui correspond à la conception traditionnelle des sciences humaines. Ces dernières se modernisent, mais à très petit pas. Les choses sont différentes pour les bibliothèques et l'édition.

Avez-vous le sentiment d'appartenir à une avant-garde?

Non. On m'a attribué ce rôle – ainsi qu'à d'autres – sans que je l'aie cherché, car je suis plutôt conservateur dans ma manière de pratiquer l'analyse littéraire. Bien entendu, dans un champ nouveau, presque tout le monde se connaît, et un esprit d'équipe émerge. J'ai des contacts aussi dans les centres suisses, à Lausanne, à Bâle et à Berne. Leur méthodologie fait aussi que les humanités numériques sont plus collaboratives que les sciences humaines ne le sont d'habitude.

«Les sciences humaines se modernisent, mais à très petits pas.»

Les humanités numériques se soucient-elles suffisamment de savoir, par exemple, dans quelle mesure l'utilisation du numérique modifie le statut du texte?

Là, il y a des déficits, car les questions fondamentales et le travail pratique n'ont souvent pas encore été réunis, alors qu'on débat beaucoup de critique d'algorithmes et du texte numérisé. Mais il ne faut pas oublier que nous n'en sommes qu'au début, que nous évoluons souvent en marge, et que nous ne disposons pas encore de méthodologie confirmée.

Les sciences humaines dans l'ensemble évoluent en marge.

Si elles continuent comme elles l'ont fait jusqu'ici, leur avenir ne sera pas brillant. Leur situation est difficile. Dans les pays anglo-saxons, elles ne reçoivent pratiquement plus de fonds publics et doivent se financer presque intégralement par les écolages.

Voyez-vous les humanités numériques comme une bouée de sauvetage?

Non. Dans le meilleur des cas, comme une partie de la solution. Indépendamment des méthodes numériques, les sciences humaines doivent répondre à une grande question: celle de savoir ce qu'elles veulent enseigner à l'avenir. Et ce sans perdre de vue les changements fondamentaux que la modernisation numérique exige de nous.

L'historien Urs Hafner est journaliste scientifique et ancien rédacteur en chef d'Horizons.

Un penseur non conventionnel

Les humanités numériques constituent un point fort des travaux de Gerhard Lauer, qui enseigne la philologie allemande à l'Université de Göttingen (D). Il a récemment publié des travaux consacrés à Thomas Mann et Joane K. Rowling, et a étudié les réactions de la pupille chez le lecteur.