Robot Journalism

When computational linguistics meets journalism

Laurence Dierickx – 2017

Definitions

Computational journalism

Combination of algorithms, data and knowledge of the social sciences to complete the responsibility function of journalism. The term "computational journalism" appears for the first time in 2006 at the Georgia Institute of Technology. Also called "algorithmic journalism”, it covers the whole journalistic process.

When computational linguistics meets journalism

Automated journalism

Automated journalism is the practice of using technology to automatically produce news content. Nicknamed “robot journalism,” it uses algorithms to generate stories…no humans necessary.
Through patterned searches, the technology finds relevant data and structures it to create a presentable piece of writing or media, including graphs, maps, charts, pictures, and videos. (Source: Wibbitz)

Also called « robot journalism »

Process based on NLG (natural language generation) technologies, subfield of AI and computational linguistics.

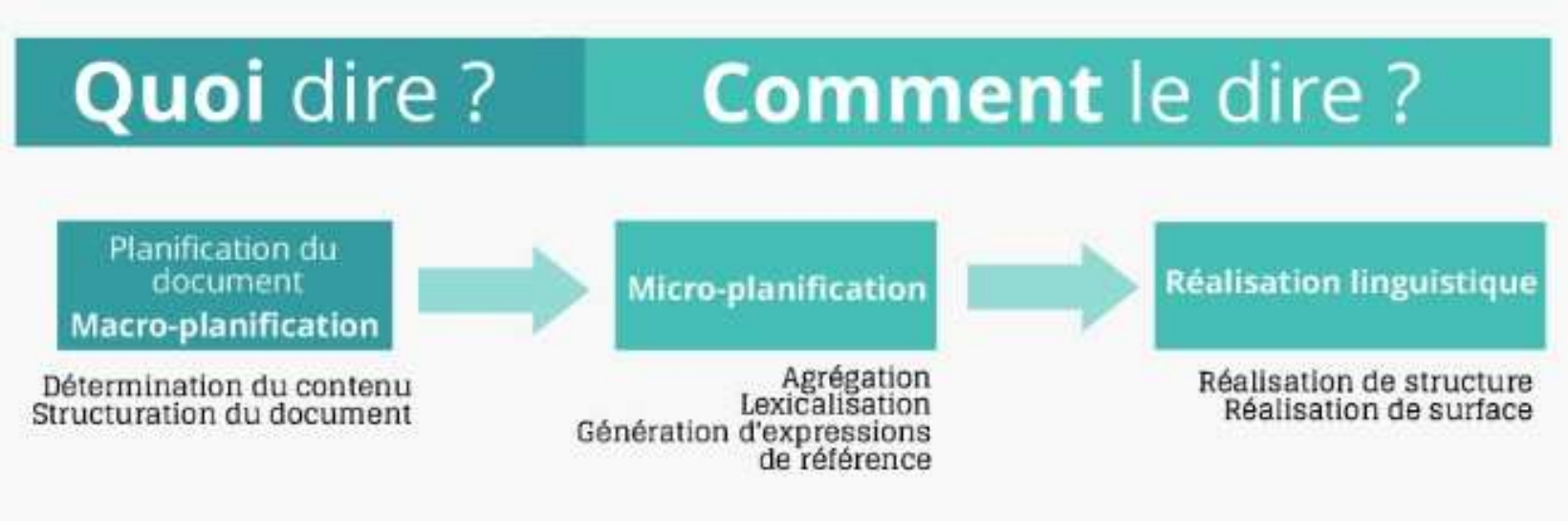
When computational linguistics meets journalism

What to say and how to say it ?

Since the beginning, computers have produced texts in natural language. Example: 'Your printer does not have any paper.' But this type of message, which is only displayed when needed after a print order, is pre-recorded and requires no 'intelligence'.
(Danlos, 1991)

Natural Language Generation (NLG) systems--computer software systems automatically generate understandable texts in English or other human languages. NLG systems use knowledge about language and the application domain to automatically produce documents, reports, explanations, help messages, and other kinds of texts.
(Reiter and Dale, 2000)

Theorical process



Pipeline architecture (Dale and Reiter)

From data (input) to text (output)

Main constraints

- data quality (for quality information)
- knowledge of the application domain (linguistics approach)

Different possible process

- Sentences with holes (basic)
- Based rules systems (if…. else…)
- Machine learning systems (related to artificial intelligence)
- Linguistics models vs stochastics models

Assets

- Accuracy (not two times the same mistake)
- Flexibility (could be easily personalized)
- Could be easily multilingual
  (working with an unique database)
- High speed level

Weakness

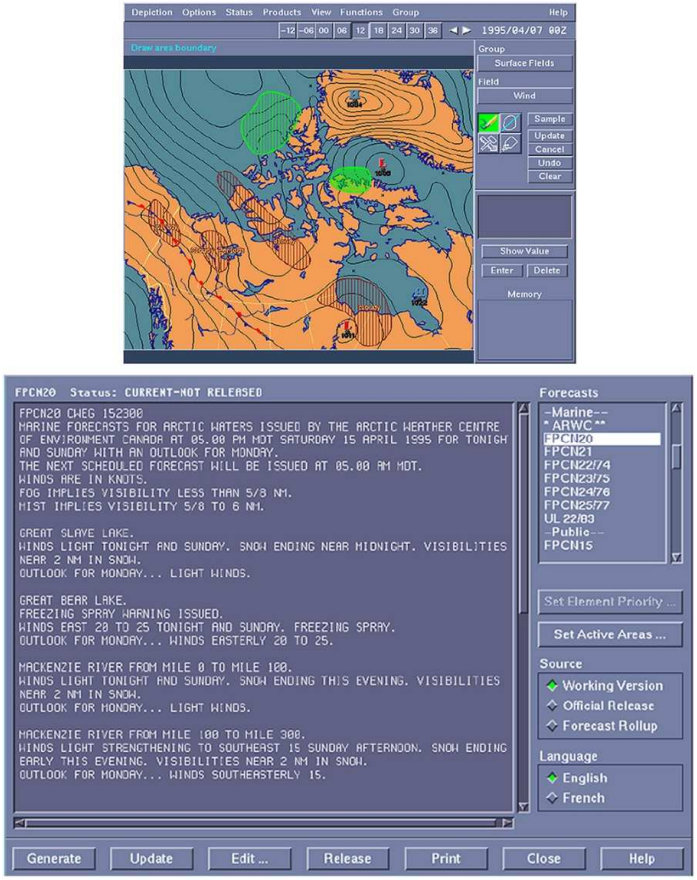
- Need high level data quality (limitation)
- Cost in time and in money too
- Standardisation vs creativity

NLG and journalism: meeting point

The whole process is ALWAYS a matter of choices just like in any editorial process

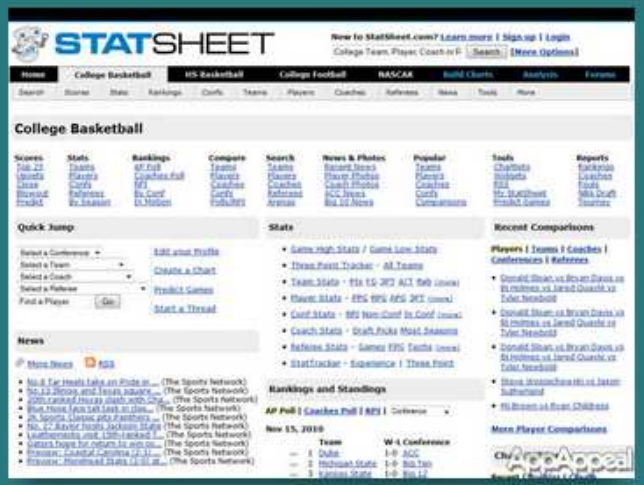
A brief history of NLG

- Field of natural language processing (NLP)
- Growing since the late 1960's : playing field for reseachers
- First operational systems at the beginning of the 1990's (weather reports/forecasts)
- Various application domains and experiences: health care, manual for products, poetry/songs, business reports...



FOG 1990

NLG and journalism: a growing history

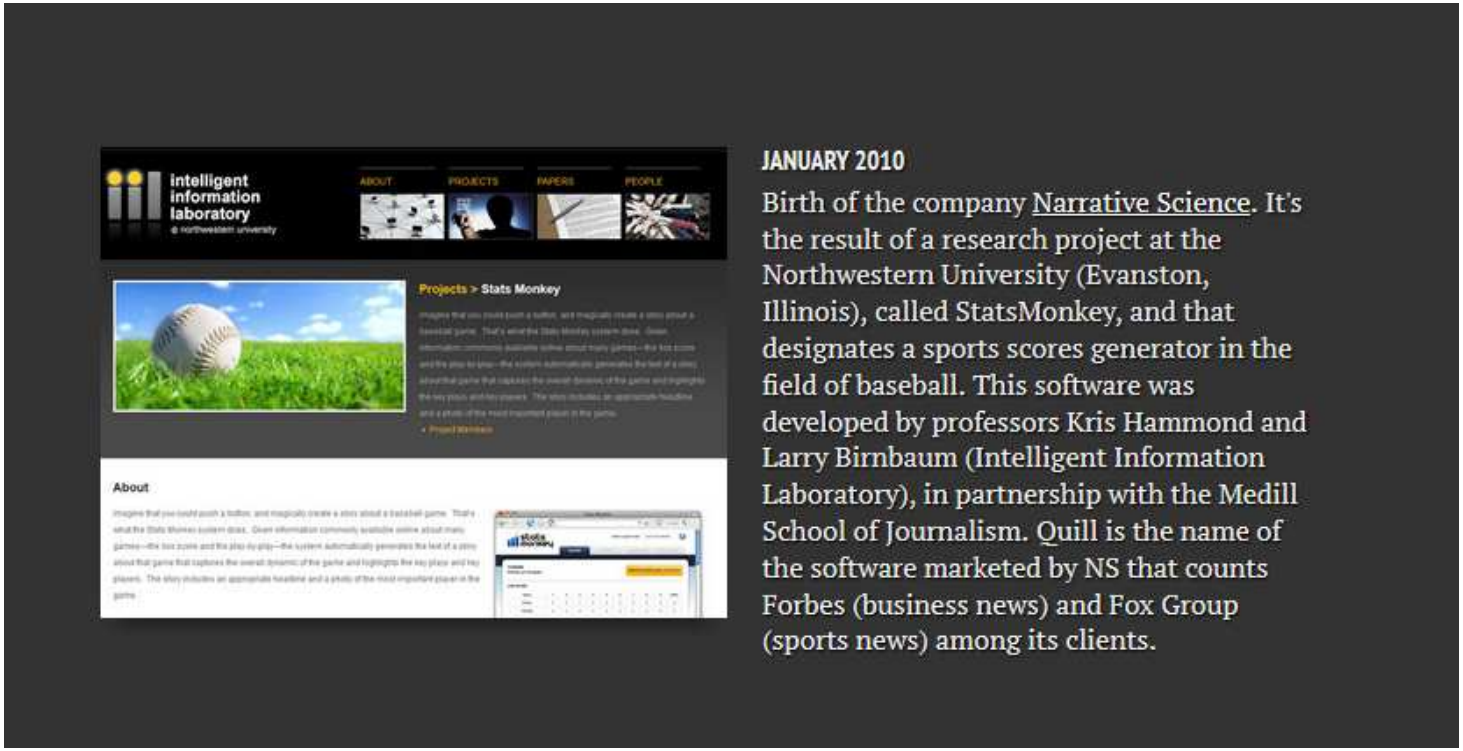


JUNE 2007

Creation of the company StatSheet in Durham (North Carolina).Its activities cover, initially, sports results. Its founder, Robbie Allen, is graduated from the Massachusetts Institute of Technology (science, engineering and management).

When computational linguistics meets journalism

NLG and journalism: a growing history



JANUARY 2010

Birth of the company Narrative Science. It's the result of a research project at the Northwestern University (Evanston, Illinois), called StatsMonkey, and that designates a sports scores generator in the field of baseball. This software was developed by professors Kris Hammond and Larry Birnbaum (Intelligent Information Laboratory), in partnership with the Medill School of Journalism. Quill is the name of the software marketed by NS that counts Forbes (business news) and Fox Group (sports news) among its clients.

When computational linguistics meets journalism

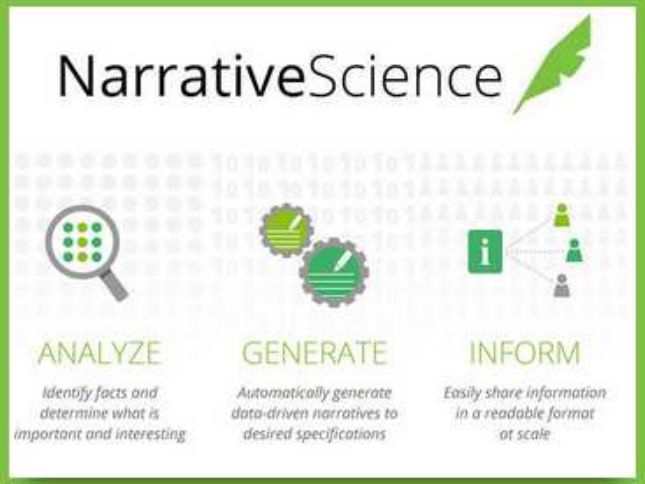
NLG and journalism: a growing history

SEPTEMBER 2011

StatSheet changes its name to become Automated Insights. AI broadens the spectrum of its activities outside the sports domain (marketing reports, financial reports...) , and raises $ 4 million at the same time.

When computational linguistics meets journalism

NLG and journalism: a growing history

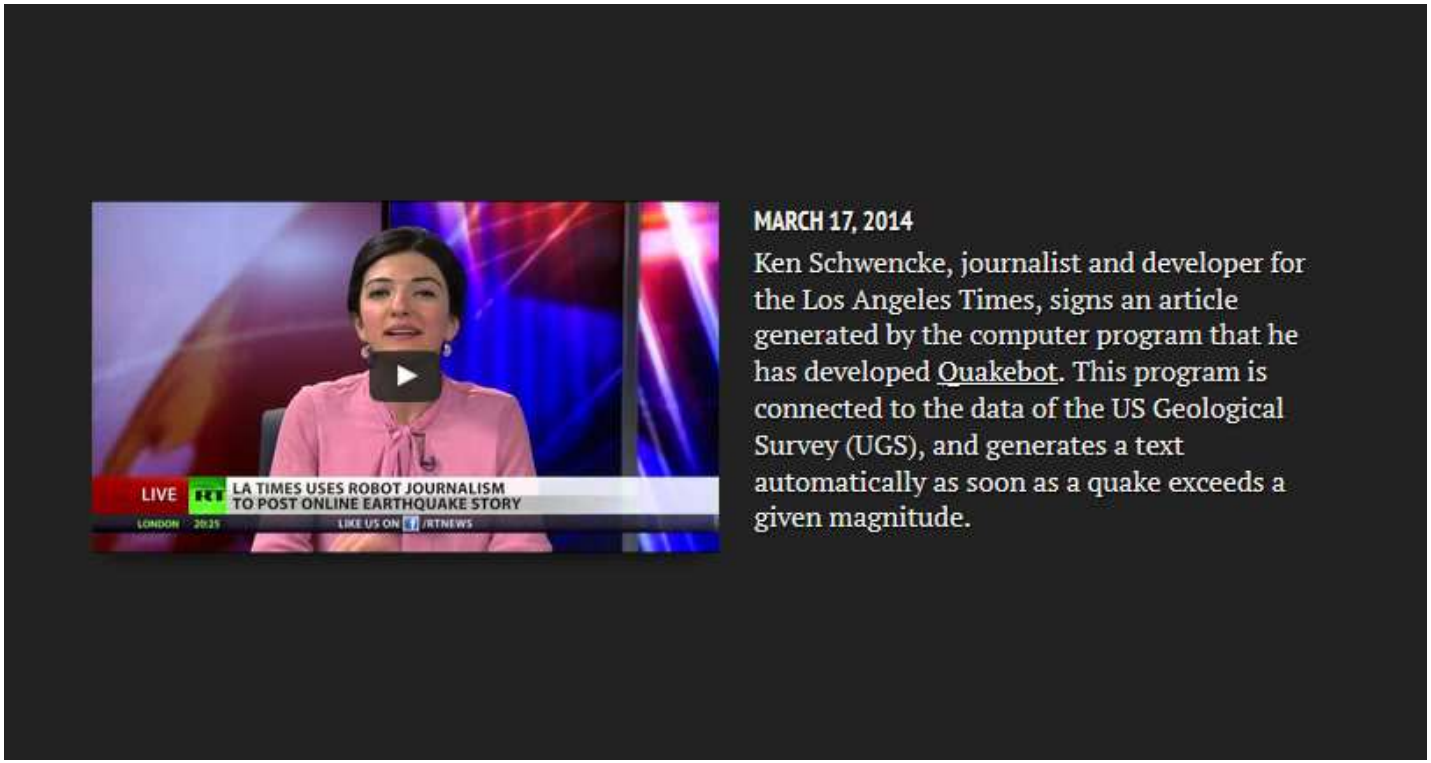


SEPTEMBER 10, 2011

Kris Hammond, co-founder of Narrative Science, says The New York Times that "in five years, a computer program will win a Pulitzer Prize — and I'll be damned if it's not our technology."

When computational linguistics meets journalism

NLG and journalism: a growing history

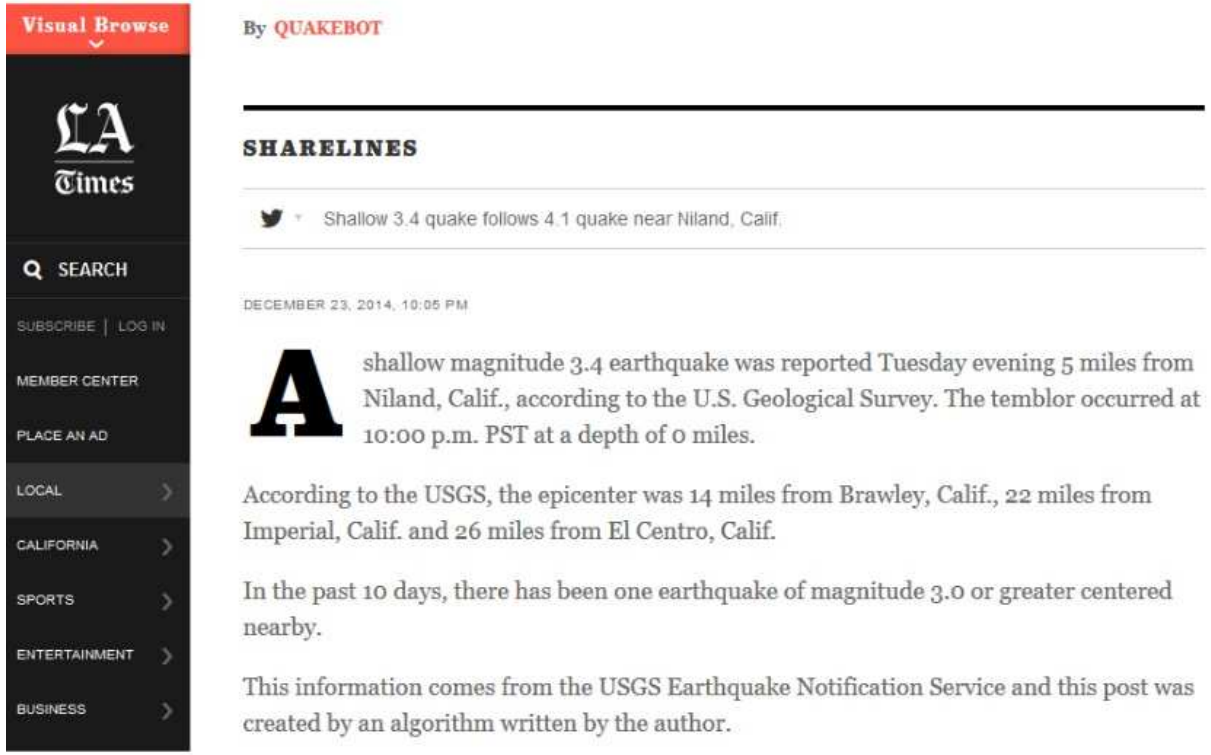


MARCH 17, 2014

Ken Schwencke, journalist and developer for the Los Angeles Times, signs an article generated by the computer program that he has developed Quakebot. This program is connected to the data of the US Geological Survey (UGS), and generates a text automatically as soon as a quake exceeds a given magnitude.

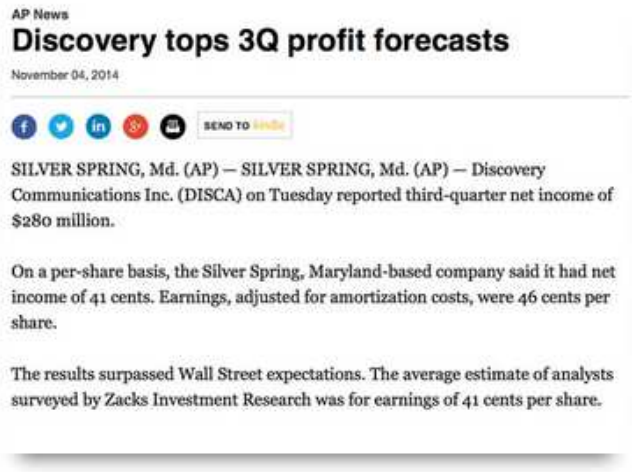
When computational linguistics meets journalism

NLG and journalism: a growing history



When computational linguistics meets journalism

NLG and journalism: a growing history



AP News

Discovery tops 3Q profit forecasts

November 04, 2014

SILVER SPRING, Md. (AP) — SILVER SPRING, Md. (AP) — Discovery Communications Inc. (DISCA) on Tuesday reported third-quarter net income of $280 million.

On a per-share basis, the Silver Spring, Maryland-based company said it had net income of 41 cents. Earnings, adjusted for amortization costs, were 46 cents per share.

The results surpassed Wall Street expectations. The average estimate of analysts surveyed by Zacks Investment Research was for earnings of 41 cents per share.

JUNE 30, 2014

Associated Press announces that the technology developed by Automated Insights (Wordsmith) will be used to produce corporate earnings stories, from with data from Zacks Investment Research. "This is about using technology to free journalists to do more journalism and less data processin", says AP. 4.300 pieces are generated by quarter, three times more than before. Gannett group (USA Today, Yahoo News) is another client of Automated Insights.

When computational linguistics meets journalism

NLG and journalism: a growing history

When computational linguistics meets journalism

NLG and journalism: a growing history

FEBRUARY 24, 2015

Tom Kent, deputy managing editor of the Associated Press and journalism instructor (Columbia University), publishes an ethical checklist for robot journalism. His recommendations focus on the accuracy of the data, the right to use them, the diversity of phrasing, the signature of the generated article or the system's governance.

NLG and journalism: a growing history

Départementales 2015

POLITIQUE ELECTIONS DÉPARTEMENTALES 2015 Elections départementales : les résultats département

France › Franche-Comté › Jura › Saint-Amour - 3913 - Résultats des élections départementales

SAINT-AMOUR 3913

POPULATION EN 2012 : 13 441 habitants PRÉSIDENT DU CONSEIL GÉNÉRAL SORTANT : Christophe PERNY

Résultats du second tour des élections départementales (29 mars 2015) : canton de Saint-Amour (Jura)

Dans la triangulaire opposant les binômes de l'Union de la Droite, à ceux de Divers gauche et du Front National dans le canton de Saint-Amour (Jura), lors du second tour des élections départementales, c'est le tandem composé de M. FRANCHI Jean et de Mme PELISSARD Hélène (Union de la Droite) qui a remporté les élections, avec 40,49 % des suffrages exprimés.

Mme BRENOT Valérie et M. FOURNIER Fernand (Divers gauche) ont été battus avec 32,01 % des suffrages exprimés, suivis de M. CAIRE Nicolas et de Mme LEGER Emy (Front National) avec 27,5 % des voix.

Dans ce canton, 37,43 % des inscrits ne se sont pas présentés aux urnes.

Ces tables ont été écrits en collaboration avec Data2Content, une marque de la société Syllabs, à partir des données du Ministère de l'Intérieur et de l'Insee.

MARCH 22, 2015

On the occasion of the first round of the French departmental elections, Le Monde begins a partnership with the French start-up Syllabs, founded in 2006, to generate reports. 30.000 articles were produced and published in one night, with the software Data2Content. This automated material have been integrated into the online platform Données du Monde.

When computational linguistics meets journalism

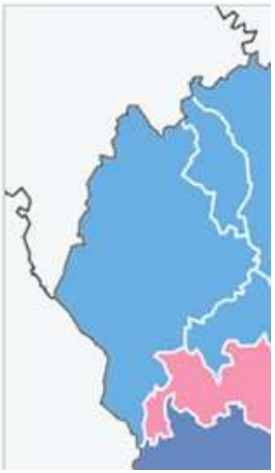
NLG and journalism: a growing history

LA CANOURGUE 4802

POPULATION EN 2012 : 6 501 habitants **PRÉSIDENT DU CONSEIL GÉNÉRAL** : Jean-Paul POURQUIER

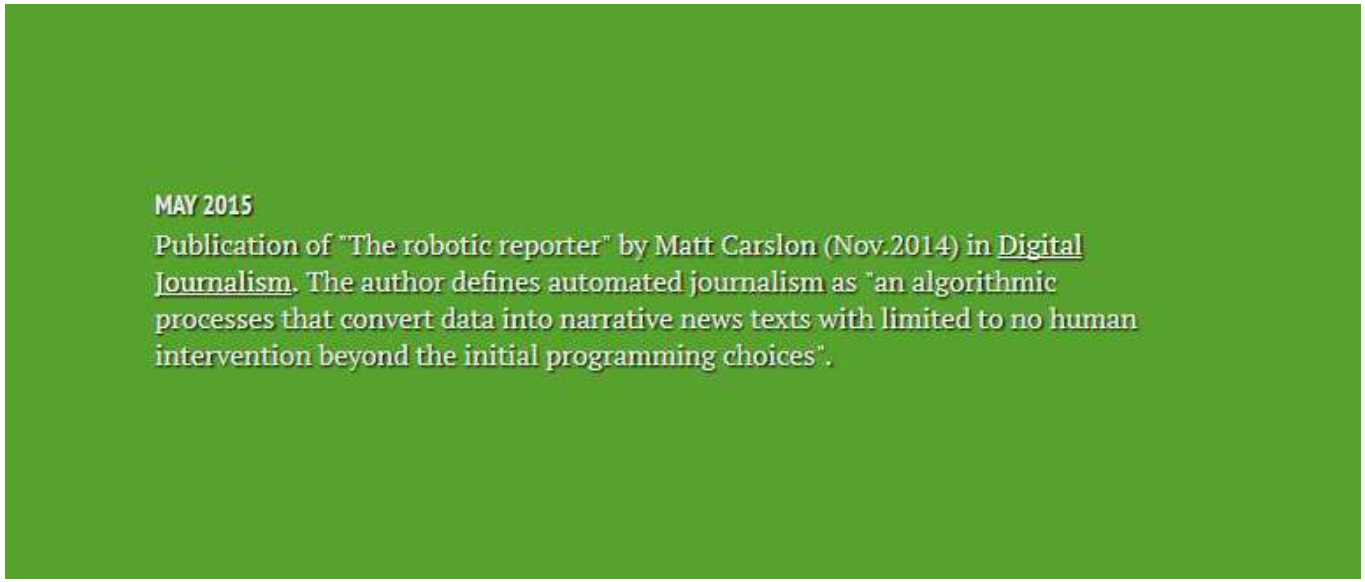
Résultats du premier tour des élections départementales (22 mars 2015) : canton de La Canourgue
Le binôme constitué de Mme FABRE Valérie et de M. POURQUIER Jean-Paul (Union pour un Mouvement Populaire) est arrivé en tête du premier tour des élections départementales, dimanche 22 mars, dans le canton de La Canourgue avec 42,69 % des suffrages exprimés. Le tandem de l'Union pour un Mouvement Populaire, devance le binôme du Modem formé par Mme AULAS Marie-Dominique et M. ROCHOUX Philippe, qui a obtenu 27,23 % des voix. Ces deux binômes sont qualifiés pour le second tour, dimanche 29 mars. Les candidats éliminés sont : Mme BONICEL Arlette et M. GAUDRY François (Front de Gauche) 16,75 % et Mme DAMIEN Marie-José et M. GARDETTE Philippe (Front National) 13,33 %. Le taux d'abstention a atteint le score de 27,63 % dans ce canton.

Ces textes ont été écrits en collaboration avec Data2Content, une marque de la société Syllabs

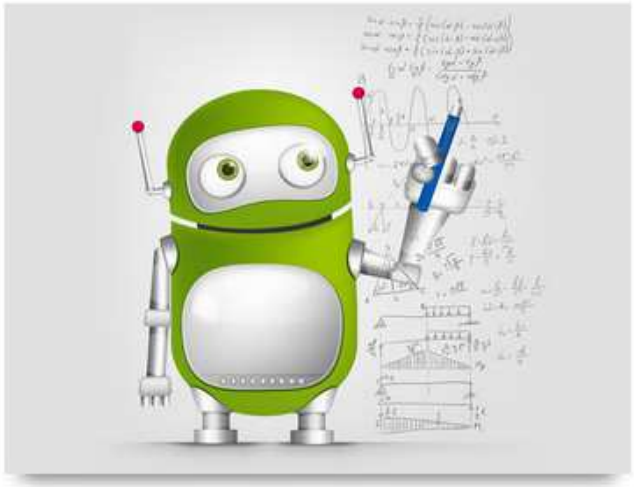


When computational linguistics meets journalism

NLG and journalism: a growing history



NLG and journalism: a growing history



MAY 2015

The Dutch government has funded research into robotic journalism with giving a grant of €700.000 to the Tilburg University. This research is supported by a the media organisation NDP (members: Persgroep, RTL, ANP). Some research will be carried out at the Telegraaf (newspaper).

When computational linguistics meets journalism

NLG and journalism: a growing history

SEPTEMBER 2015

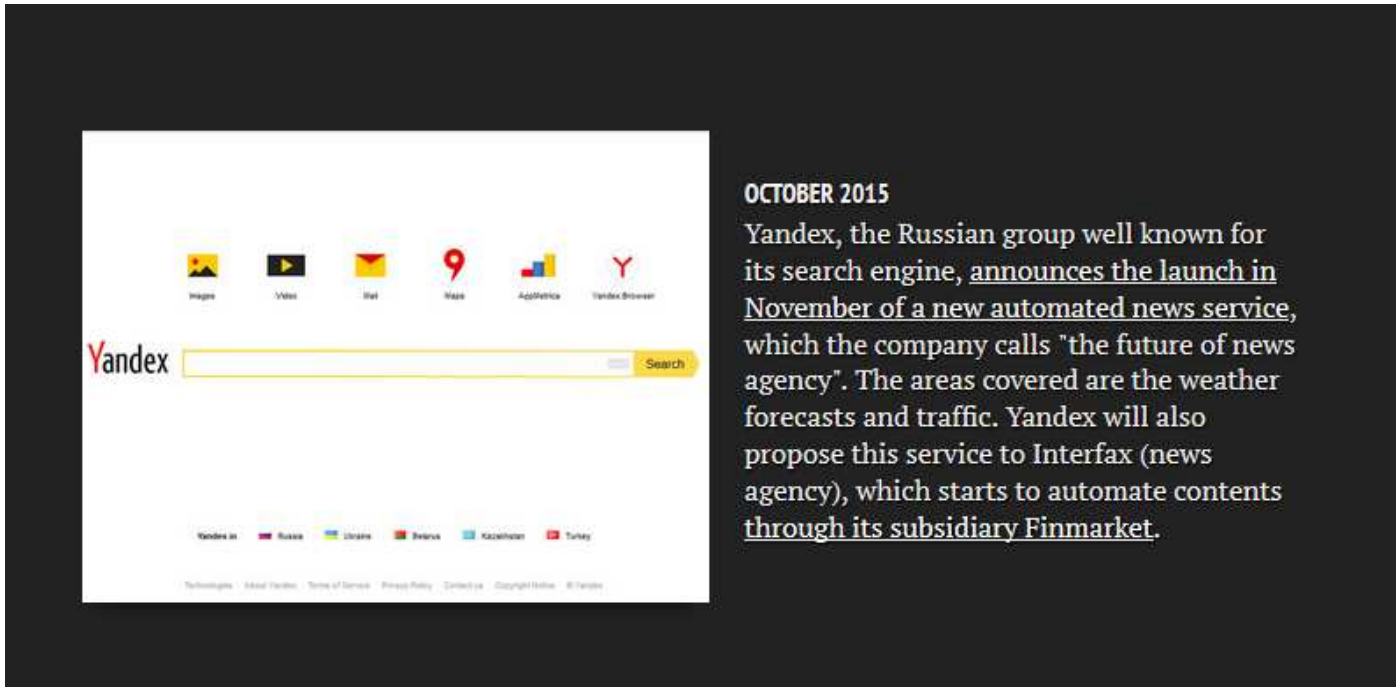
The local publisher On The Wight, has built a program (in Python) to automate unemployment reports. More data sets could be cover in the future.

When computational linguistics meets journalism

NLG and journalism: a growing history

When computational linguistics meets journalism

NLG and journalism: a growing history



When computational linguistics meets journalism

NLG and journalism: a growing history

OCTOBER 2015

MittMedia, a Swedish media group "hired" two "robots" journalists. The first one writes 45 weather reports every morning for the various titles of the group; while the second aims to communicate with journalists about page views on mobile devices. Next step: automate reports of football matches.

When computational linguistics meets journalism

NLG and journalism: a growing history

OCTOBER 20, 2015

Automated Insights launches the beta version of its software Wordsmith. Users don't need to know coding or having an experience with data science to create personalized stories, articles and reports. The software, powered by a CSV file (spreadsheet) proposes a story structure to be enriched. Once created, the model can be used for an unlimited number of articles.

When computational linguistics meets journalism

NLG and journalism: a growing history

When computational linguistics meets journalism

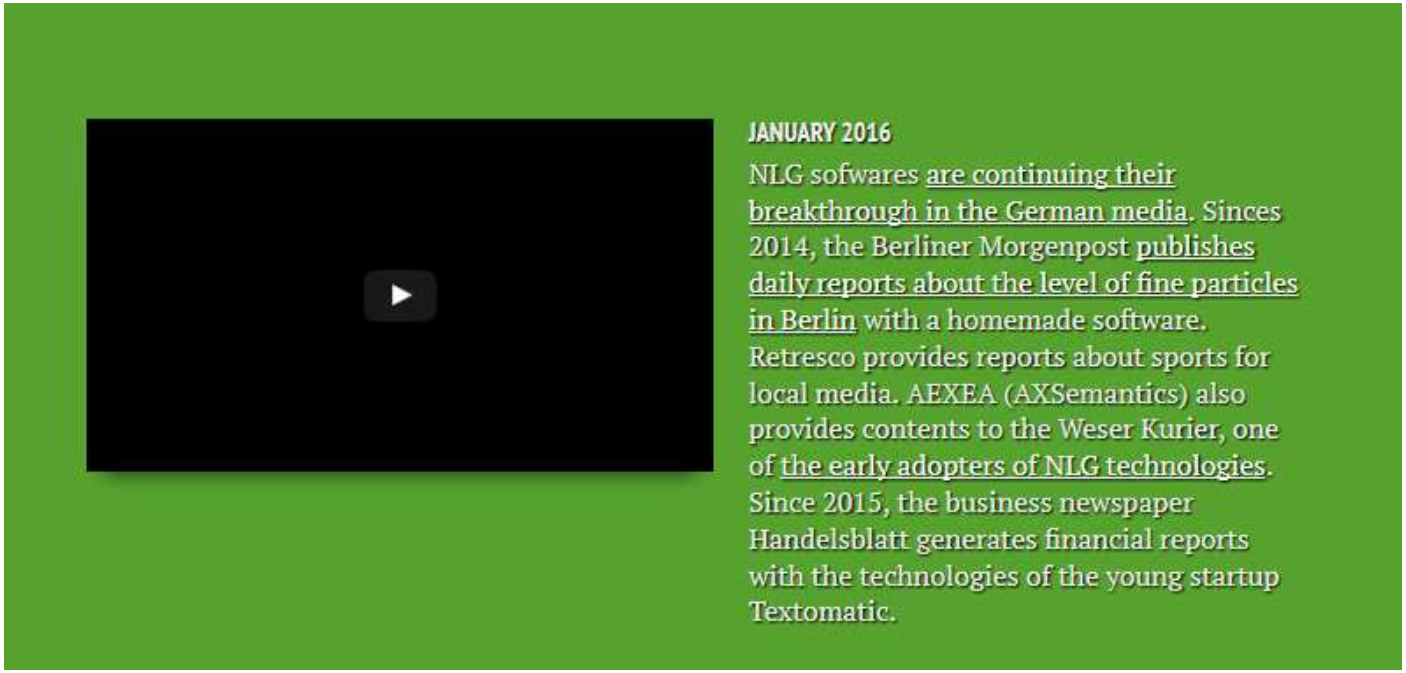
NLG and journalism: a growing history

When computational linguistics meets journalism

NLG and journalism: a growing history

When computational linguistics meets journalism

NLG and journalism: a growing history

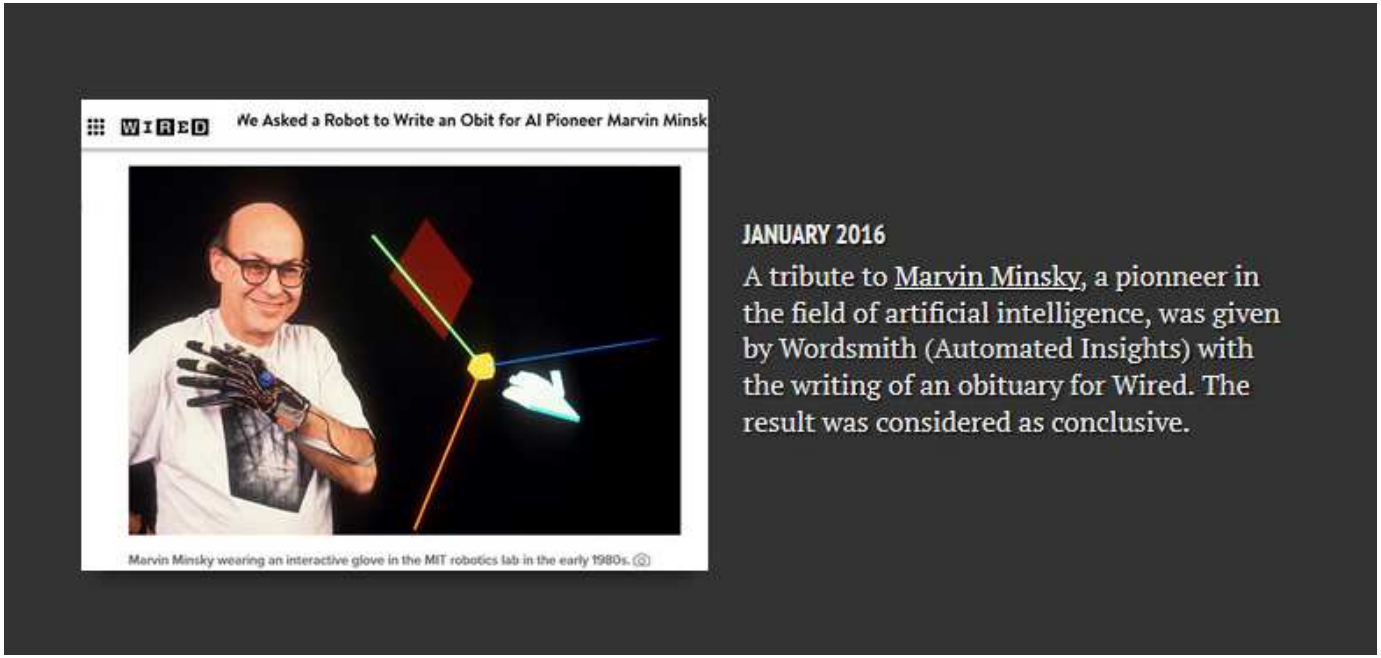


JANUARY 2016

NLG sofwares are continuing their breakthrough in the German media. Sinces 2014, the Berliner Morgenpost publishes daily reports about the level of fine particles in Berlin with a homemade software. Retresco provides reports about sports for local media. AEXEA (AXSemantics) also provides contents to the Weser Kurier, one of the early adopters of NLG technologies. Since 2015, the business newspaper Handelsblatt generates financial reports with the technologies of the young startup Textomatic.

When computational linguistics meets journalism

NLG and journalism: a growing history



When computational linguistics meets journalism

NLG and journalism: a growing history

Image: CanStockPhoto

MARCH 2016

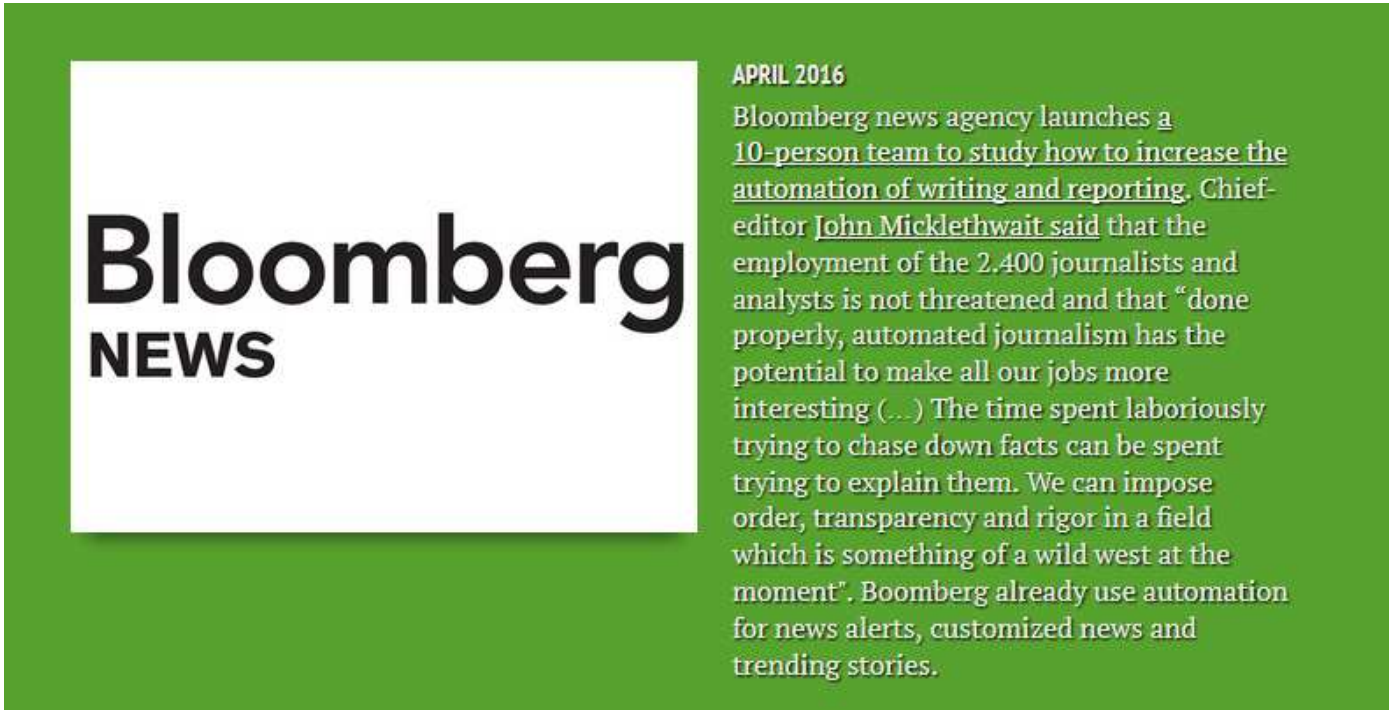
In Sweden, Östgöta Media in collaboration with the company UNT has developed Rosalinda, a bot that writes about football matches. Objective: to ensure a better coverage of local events.

When computational linguistics meets journalism

NLG and journalism: a growing history

When computational linguistics meets journalism

NLG and journalism: a growing history



When computational linguistics meets journalism

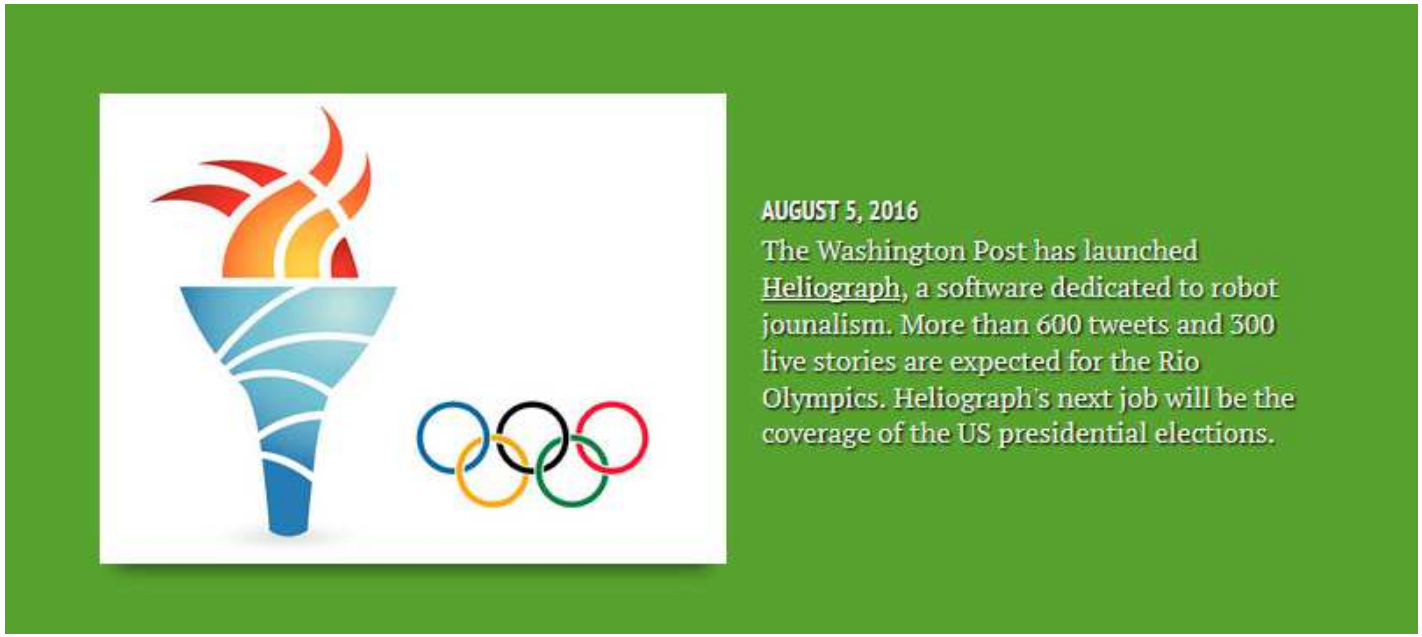
NLG and journalism: a growing history

JUNE 2016

Adrienne Lafrance, journalist at The Atlantic, have experiment a software trained to write like her with the use of a neural network, a computer model inspired by the human brain. Robot Adrienne had to learn from articles published by the journalist since 2014. About 725,000 words were not enough to achieve a satisfactory result: "For now, machine journalists should probably stick to box scores and basic weather reports."

When computational linguistics meets journalism

NLG and journalism: a growing history



AUGUST 5, 2016

The Washington Post has launched Heliograph, a software dedicated to robot jounalism. More than 600 tweets and 300 live stories are expected for the Rio Olympics. Heliograph's next job will be the coverage of the US presidential elections.

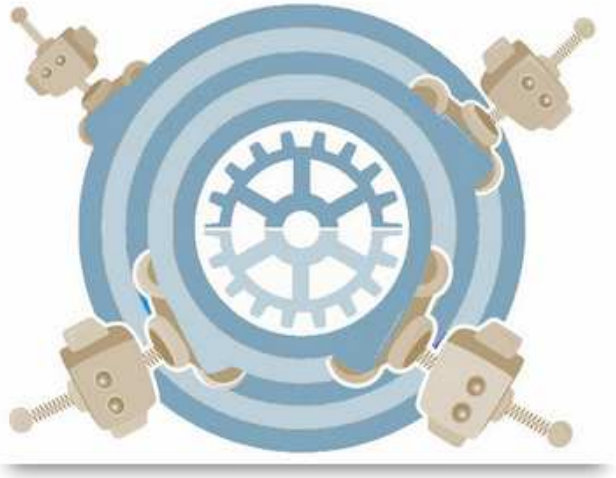
When computational linguistics meets journalism

NLG and journalism: a growing history

SEPTEMBER 2016

In Spain, Narrativa proposes automated reporting about sports. The start-up aims to cover minor league soccer games from other countries than Spain.

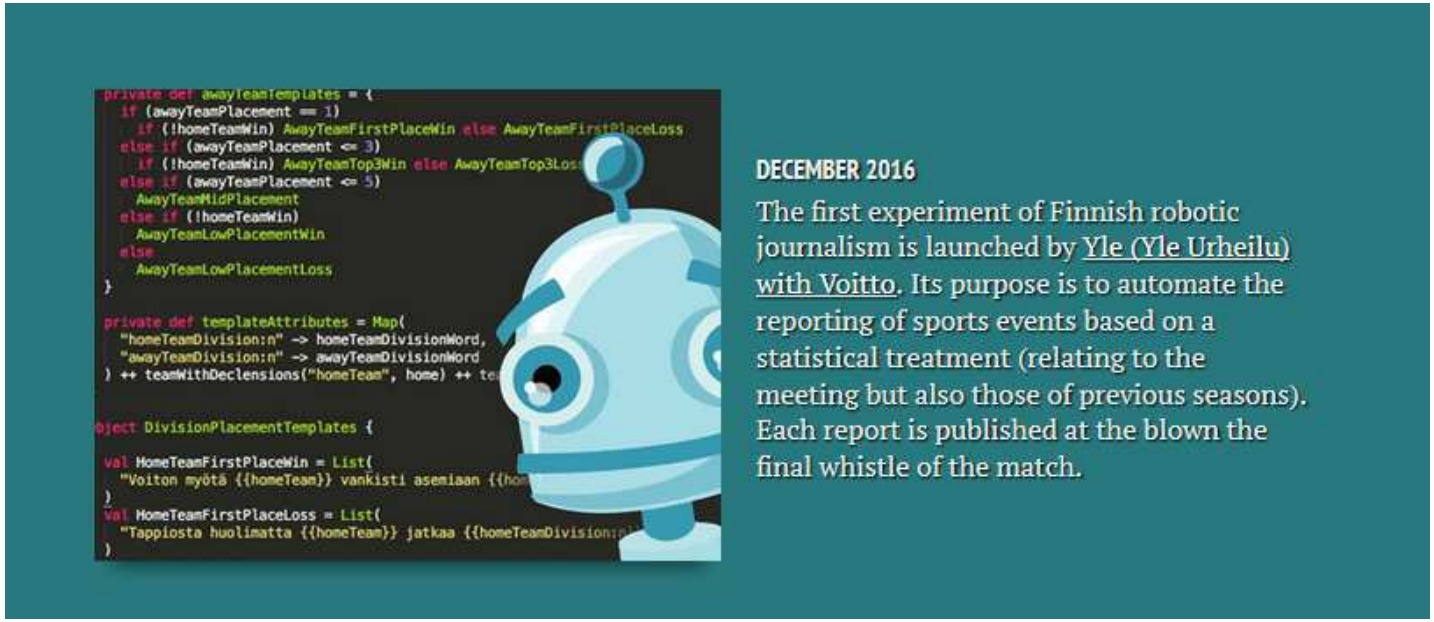
NLG and journalism: a growing history



OCTOBER 2016

Press Association has announced to start looking at using automated journalism for business, sport and election coverage in the next few months. According to Pete Clifton, editor-in-chief of the news agency, the purpose is not to replace journalists ("This won't be replacing any of our fantastic journalists") but to provide "an extra level when it comes to short market reports, election results and football reporting".

NLG and journalism: a growing history



When computational linguistics meets journalism

NLG and journalism: a growing history


大都会 09

南都机器人小南播报：

20日广州—武汉还有大量无座票

广东半月即报告11例H7N9病例

发病水平高于2016年同期，专家称疫情可防可控

JANUARY 2017

The adventure continues in China where the Southern Metropolis Daily published its first report written by a software. The story counted 300 mots, and "summarizes what train tickets are most in demand over the Lunar New Year holiday".

When computational linguistics meets journalism

NLG is only one of the aspects of automation…

- Automated Twitter newsfeeds, breaking news detection or fact-checking
- Newsbot messengers (Fusion for Facebook messenger, that send top headlines using emojis instead of select phrases) and other type of messenger bots (BBC World Service has laucnhed a bot on Telegram to propose a digest to Uzbekistan readers as BBC is blocked in their country) or chat bots (automated interactivity with readers) as well as robots live bloguers (The Telegraph, UK)
- In UK, Perspect proposes 3 versions of a same story (neutral, positive and negative)
- The Israelian start-up Wibbitz has launched a service that converts texts to video
- In USA, AP wants to turn print stories into broadcast ones
- Reuters has developed a suite of tools internally, to automate the process of gathering news and data with the aim to help journalists in their tasks: f. ex. „Live Data“ gathers real time information, performs calculations and „puts numbers intro prose“
- Reuters has launched a software that turn texts into data vizualisations

…and there are also (almost) „real“ robot journalists

Weather reporter in China
(Dragon TV, 2015)

In Japan: AI project research at the Tokyo University (2010). The « RJ » is able to detect changes, to take pictures, to ask questions, to publish online…

…and there are also (almost) „real“ robot journalists

First android newscaster in Japan (2014)

Robot journalism in Europe

Media : press agencies, online news media (broadcasters, daily press, pure players, magazines…)

Types of data: elections, sports, business, employment, services

Limitation: NLG softwares need good structured data in input. For this reason, covered domains are so far limited.

Main assets

- Gives time to journalists to go back tot he roots of their job / Free up time for valuable work
- Could provide different versions of a same story easily, even in different languages
- Creating content on a massive scale (1 million articles produced in one night by Data-2-Content, the software developed by Syllabs)
- High accuracy level
- Not two times the same mistake
- Extend the media coverage / „niche“ information

Main issues

- Rely on good structured data (limitations)
- Need the knowledge of an application domain, robot must be trained
- Cost: *Le Monde* has paid between €20.000 and €50.0000 for the coverage of regional election
- Impact on journalism and on journalists not yet measured BUT it is often claimed that it will contribute to create new jobs with specific skills and contribute to the evolution of journalism

Three kind of actors

1) Start-ups

Syllabs (France), AX Semantics (Germany), Textomatic (Germany), Narrativa (Spain)….

Most of those companies do not employ journalists but computational linguists, computer scientists, data scientists…

Media are clients among others, that means that those companies are also active in other fields (communication, marketing, business…) There are others NLG companies in Europe (F.E. Yseop in France, Arria NLG in UK) but they do not work for media organizations.

Those companies are not claiming themselves as news media organizations.

Three kind of actors

2) Media organizations

Most of the time : press agencies (internal developments)

Interesting case: « On the Wight », that provides local news about Isle of Wight. For the two journalists, who have hired a computer scientist, it is a way to extend their coverage while their means/resources are limited.

Three kind of actors

3) Journalists-programmers

Despite the fact that those profiles do exist in the USA (f.e. Ken Schwencke who has developed « Quakebot »), it is not so widespread and, for now, no comparable experience has been observed in Europe.

Why ? Combination of 2 types of work : journalist and developer/programmer. Between those two jobs, many differences BUT mixed together they are totally complementary.

In USA: difficulty to attract journalists to computer science or to data science but the inverse path is true: computer students are also trained in journalism in several universities programs (see further)

Ethical considerations

« Stupid piece of code » (Claude de Loupy, Syllabs): the human input remains important
→ journalistic ethical standards for all

Ethics must be applied to all of the editorial choices, including algorithmic ones

Lack of transparency: confidential disclosures disallow companies to reveal the name of their media clients: in Germany, readers are reading automated contents without even knowing it
→ being clear and honest with the readers AND with the journalists

How to build confidence with the audiences when the truth is hidden?

Rights to use or reuse the data (just like in any datajournalism project), questioning the source

Ethical considerations

The danger of personalized generated contents: it has not happened for now
BUT some voices are warning, especially in the US, about the temptation to provide contents tailor made for readers. One of the main function of journalism is to explain the world in its whole complexity. How achieve this goal if readers are enclosed in their own cultural sphere?

The temptation could more concern new online players than news media organizations.

Journalism is a profession: how to include journalists to the whole process? Might be a warrantee for an ethical information.

Ethical considerations

Tom Kent, the AP's standards editor, thinks concerns about algorithmic transparency are overblown when it comes to automatically generating content. "Human journalism isn't all that transparent," he says. "News organizations do not accompany their articles with a whole description of what was on the journalist's mind that could have affected his thinking process, whether he had a head cold, had just been hung up on by a customer service rep of the company he was writing about, and so on."

Because the rules governing how automated stories get assembled are available for scrutiny, automated journalism may be more transparent than stories written by humans (Kent).

Source : Nieman Report, Automation in the newsrooms, 2015

How could journalists deal with robot journalism?

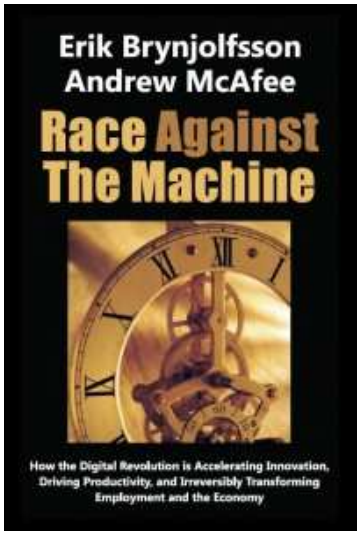
The threatened jobs don't have value-added. Robots will do the most menial tasks and will free journalists for doing things much more interesting and more rewarding.
Eric Scherer, France Télévisions

Beyond the increase in productivity, these technologies can provide legible contents on topics where there are too few readers for the work of a journalist.
Nicolas Kayser-Bril, Journalism++

It would be dangerous to resist the revolution: it is better to accompany the phenomenon to benefit from it rather than suffer. Journalists must be involved in this process.
Ricardo Gutiérrez, European Federation of Journalists

An ally rather than an adversary

Journalists can not compete with the machines: if it is a race against the machines, humans have already lose. According to Brynjolffson and McAfee, the best way is to make the machines an alley rather than adversaries.



How? To support quality investigative journalism, with being complementary

Example: news bot about air quality measurements in Brussels (real time open data), the robot gets and stores the data, provides computations, and run alone during one year. At the end of the experience, analysis are put in context and explained by journalists, a deeper work with perspectives that the bot can't do.

Building a bridge between journalists and computation

How to build a bridge? The question of the required skills belongs to universities and journalism schools.

Specific programs (as it exists for audio-visual journalism, f.e.) should be developed to professionalize journalists in this field and make them able to deal with new technologies developments as well as to be able to talk the same language as computer/data scientists.

Continuous training for journalists should be oriented to project management and PM methodologies used in the computational field (AGILE, SCRUM…)

Specific trainings on data and linguistics computation should be organized with universities, with specific programs with computer and data scientists. A one-year program should be a good starting point to get a good professionalized level.

Data and computation in journalism schools (USA)

Study published in 2016

4 key domains: data reporting, data visualisation, interactive applications, computational journalism

Based on interviews of more than 50 journalists, educators and students

The authors of this report believe that all journalism schools must broaden their curricula to emphasize data and computational practices as foundational skills

Beyond teaching, too few journalism schools support faculty research into tools and techniques of data-driven reporting, despite rich opportunities for developing theories and applications that may change journalistic practice.

Data and computation in journalism schools (USA)

Many journalism programs offer few courses in data journalism, and nearly half offer no classes at all.

The classes offered are largely introductory, and the need is still largely for the basics, such as knowing how to use a spreadsheet, understand descriptive statistics, negotiate for data, and clean a messy data set and then “interview” it to find a story.

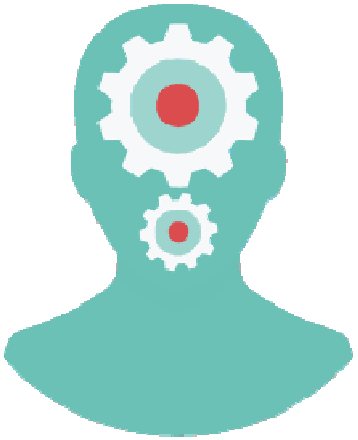
Many journalism programs do not have a faculty member skilled in data journalism.

Graduates with data journalism skills are better equipped to succeed.

Read more: https://www.gitbook.com/book/columbiajournalism/teaching-data-computational-journalism/details

New profiles are now required but it is not yet the main trend

- The amount of available data are never been so huge
- NLG is only one of the aspects about how to tell stories with data
- Automation is a key answer to deal with those massive volumes: that requires also skills in data mining, data cleaning, data analysis, data visualisations and computing.
- An example: published by The Guardian in 2015 for a data journalist position



Skills & behaviours • Skilled with desktop spreadsheet and database software, such as MS Excel and MS Access. • Some experience with server-based database platforms like SQL Server, MySQL, Oracle or Postgres. • Knowledge of basic statistics and statistical software packages (SPSS, R, Stata, etc.) is a big plus. • Basic familiarity with web technologies (HTML, CSS, Javascript, Python, Ruby, etc.) and desire to learn code. • Experience with at least one programming language is a plus. • Flexible • Self Starter • Initiative • Calm under pressure	**ALL**	

This job description is a guide to the work you will initially be required to undertake. It summarises the main aspects of the job but does not cover all the duties that the job holder may have to perform. It may be changed from time to time to meet changing circumstances. It does not form part of your contract of employment and as your experience grows, you will be expected to broaden your tasks, suggest improvements, solve problems and enhance the effectiveness of the role

Table 1 of 1
TABLE 1 How human journalists see their skills and future in the light of automated content creation

Strengths	Weaknesses
• Creativity allows human journalists to go beyond clichés and add humour to their stories • Flexibility allows human journalists to cover non-routine stories like breaking news • Analytical skills allow human journalists to go beyond description and to provide in-depth coverage	• Human journalists have higher marginal costs than automated content creation • Human journalists cannot provide the same breadth of coverage as automated content creation • Human journalists cannot compete on speed with automated content creation
Opportunities	**Threats**
• Automating routine stories gives journalists more time for research and in-depth reporting • Competition from automated content creation pushes human journalists to do a better job • Automated content creation can help to cover stories for small audiences which now go uncovered	• Automated content creation may put journalists doing routine tasks out of work • Automated content creation can be applied beyond sports reporting and also challenge the jobs of journalists in finance or real estate • Automated content creation raises new ethical questions, including issues like transparency and copyright
Source: based on 68 blog posts and newspaper articles which discuss automated sports coverage by the Statsheet Network.	

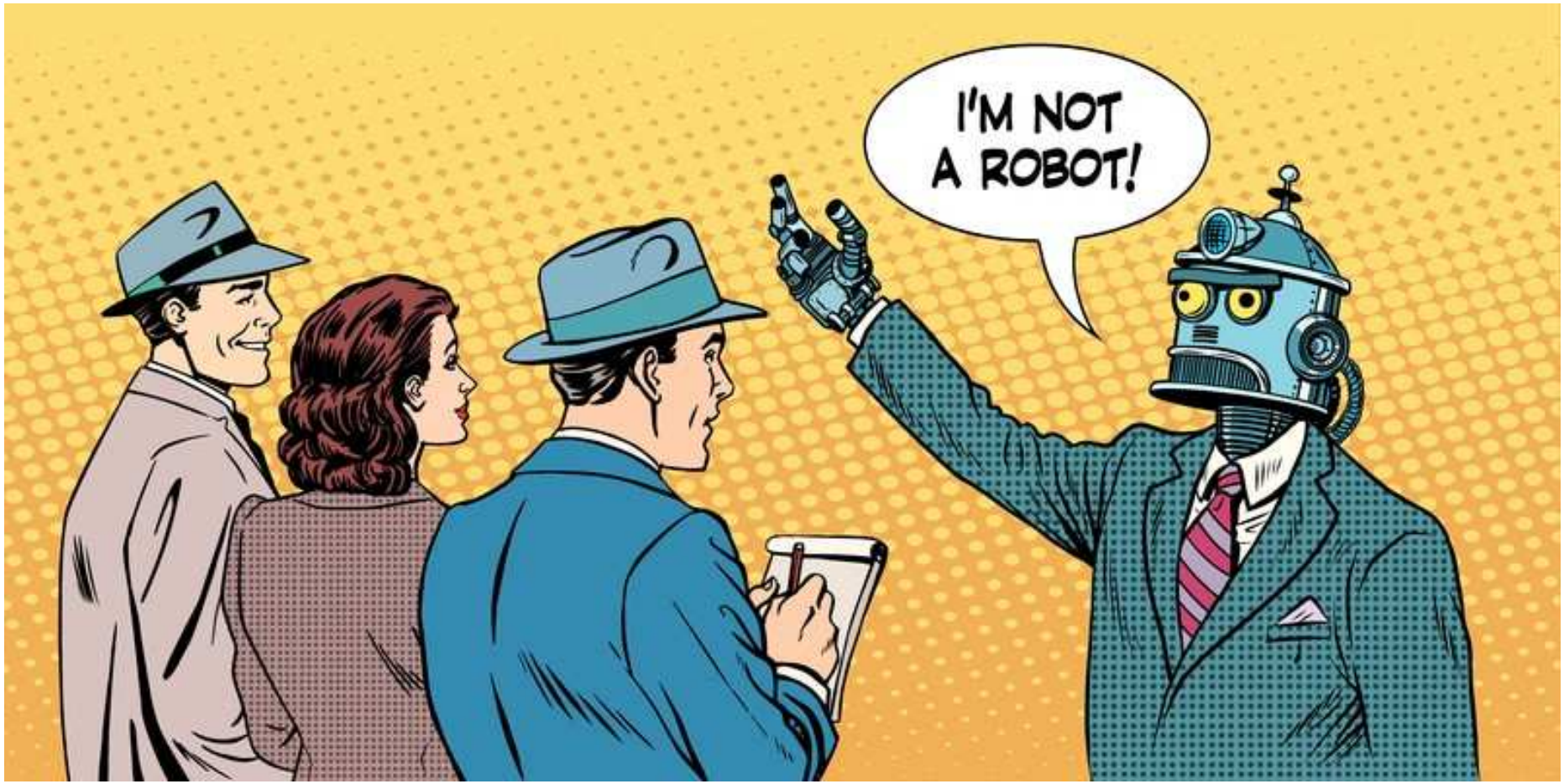
Source : Algorithms behind the headlines, Van Dalen 2012

Journalism must evolved with the digital technologies but...

It is not necessary concerning all journalists : we need deep analysis, good reports, opinions, investigative,... All those essential things that a robot can't do!

The state of the economical situation could let us fear massive unemployment due to automation. **BUT** recent studies have shown that journalism will not be the most affected compared with other profession (in the bank/insurance sector f.e). It is estimated between 8 and 12%.

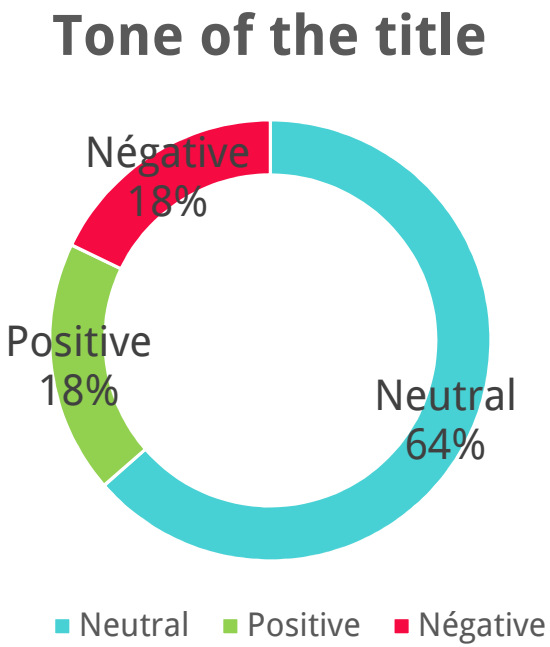
Until now, it is not yet proved that journalists have lost their jobs because of the NLG technologies, **BUT** in France, few days before the first round of the regional elections, local correspondents have learned that a robot would replace them. It was perceived as brutal. A French colleague explained that it is also an editorial point of view: the newspaper is leaving more and more the field of regional information. He said that if it is a loss for the freelancers, the main impact was observed on the relationship between journalists and their sources in terms of professional credibility.

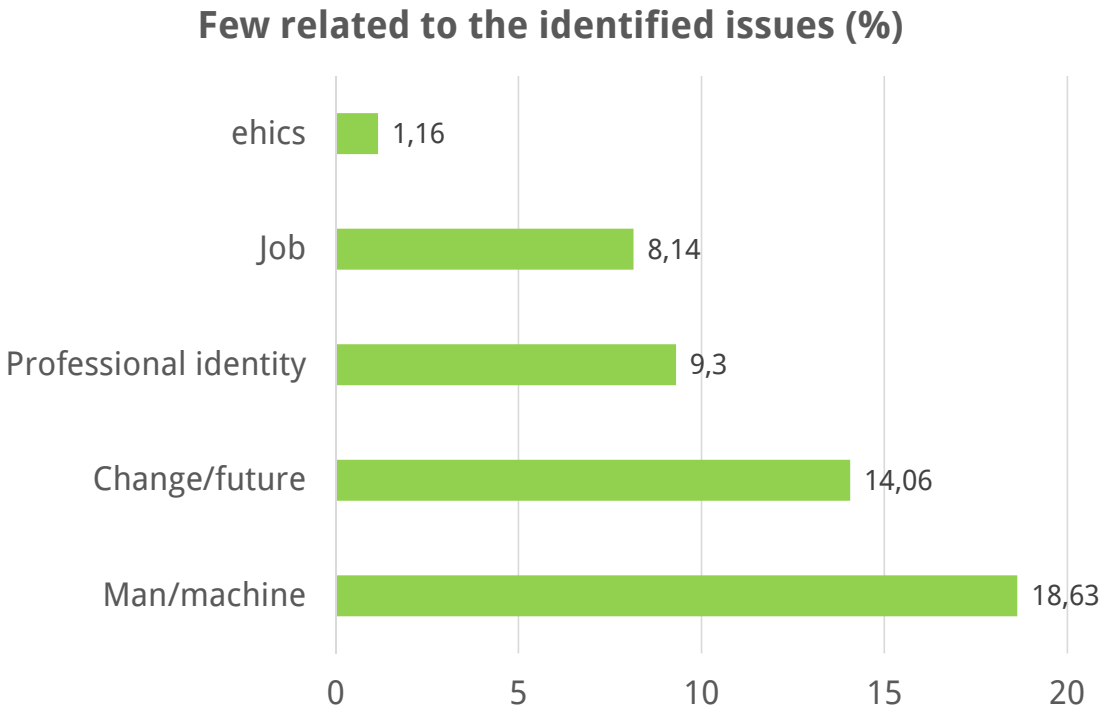


Perception of automated news articles

How do journalists talk about RJ?

Analyse of the titles of 206 articles published online between 2010 and 2016 (EN + FR)

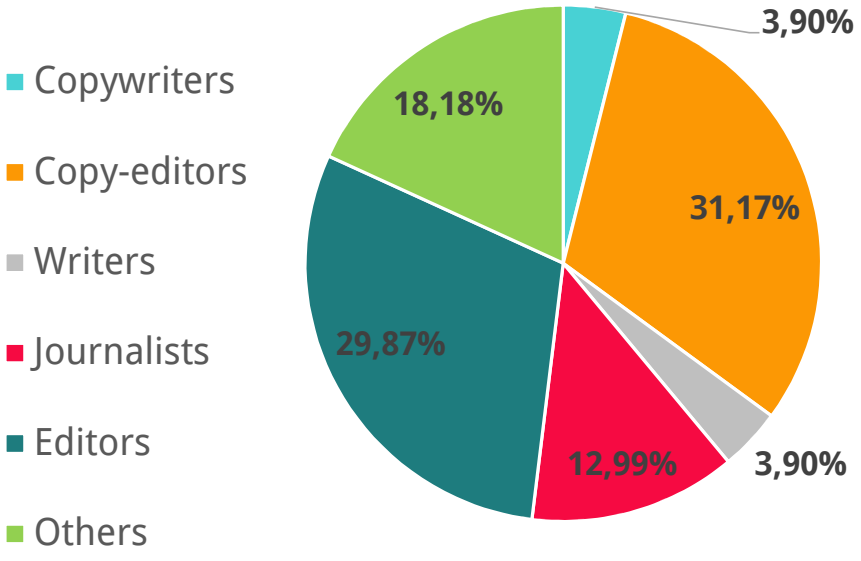




Perception of automated news articles

What do journalists think about RJ?

Research conducted in 2015 (ULB, MASTIC, Digital Information)



respondents = experts in writing (media field)

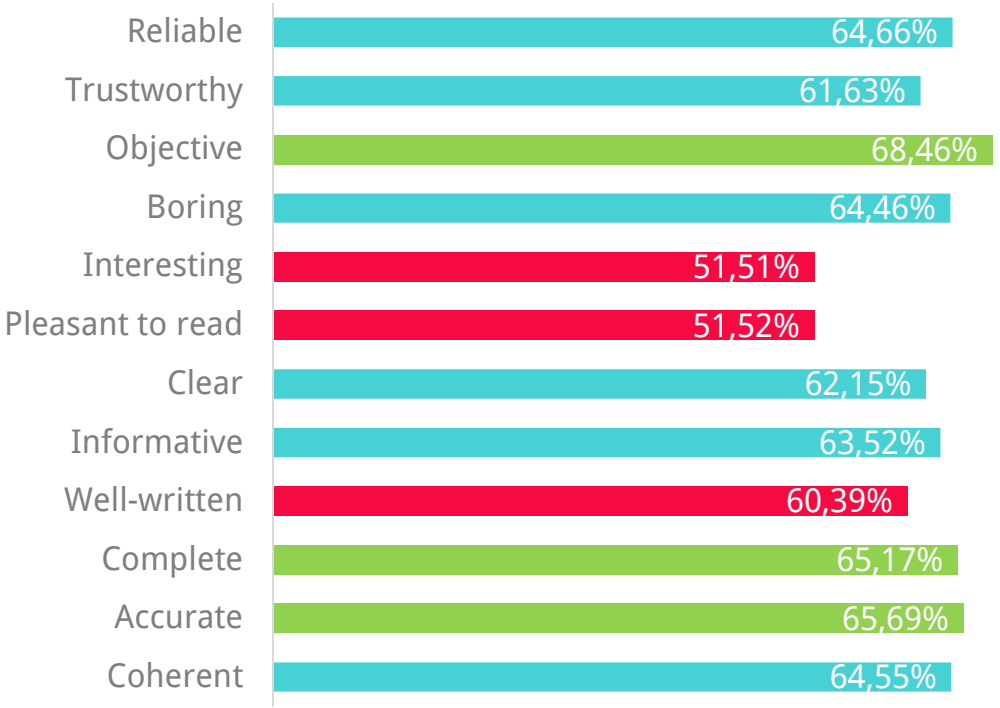
20 automated news articles first assessed with computational linguistic tools (metrics)

Automated contents get best readability scores compared to articles written by journalists (69,2% vs 62,6%)

3 articles with best scores recorded then submitted to human judges (N = 77)

Perception of automated news articles

What do journalists think about RJ?

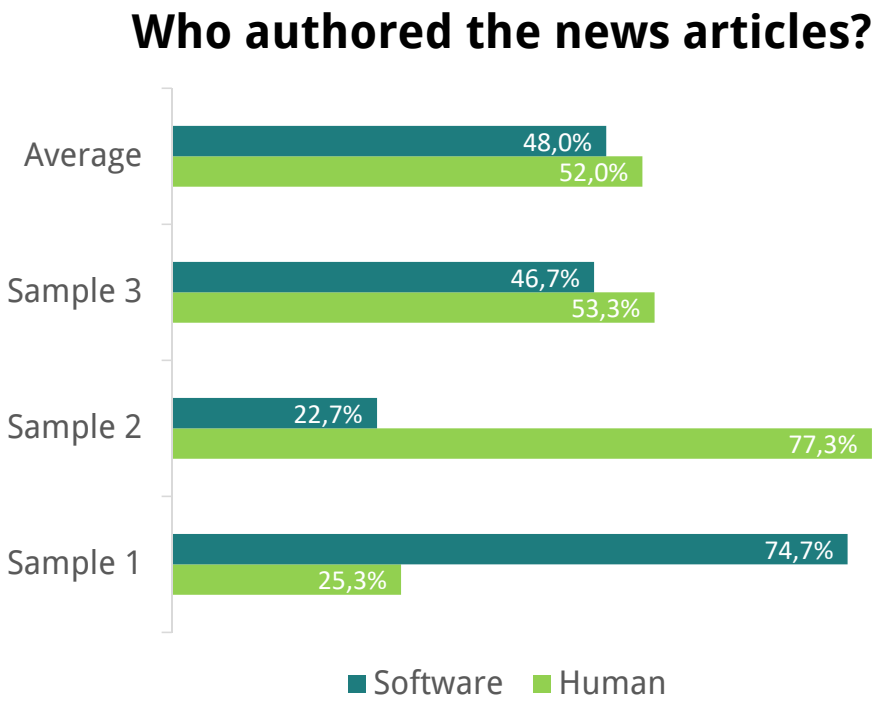


Articles about economics

Judges did not know the real subject of the experience (assessment of the quality of articles on economics published online)

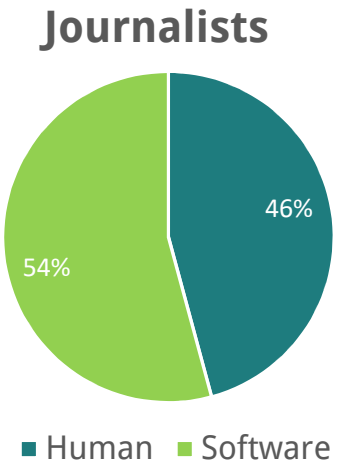
Perception of automated news articles

Were they able to recognize the software?

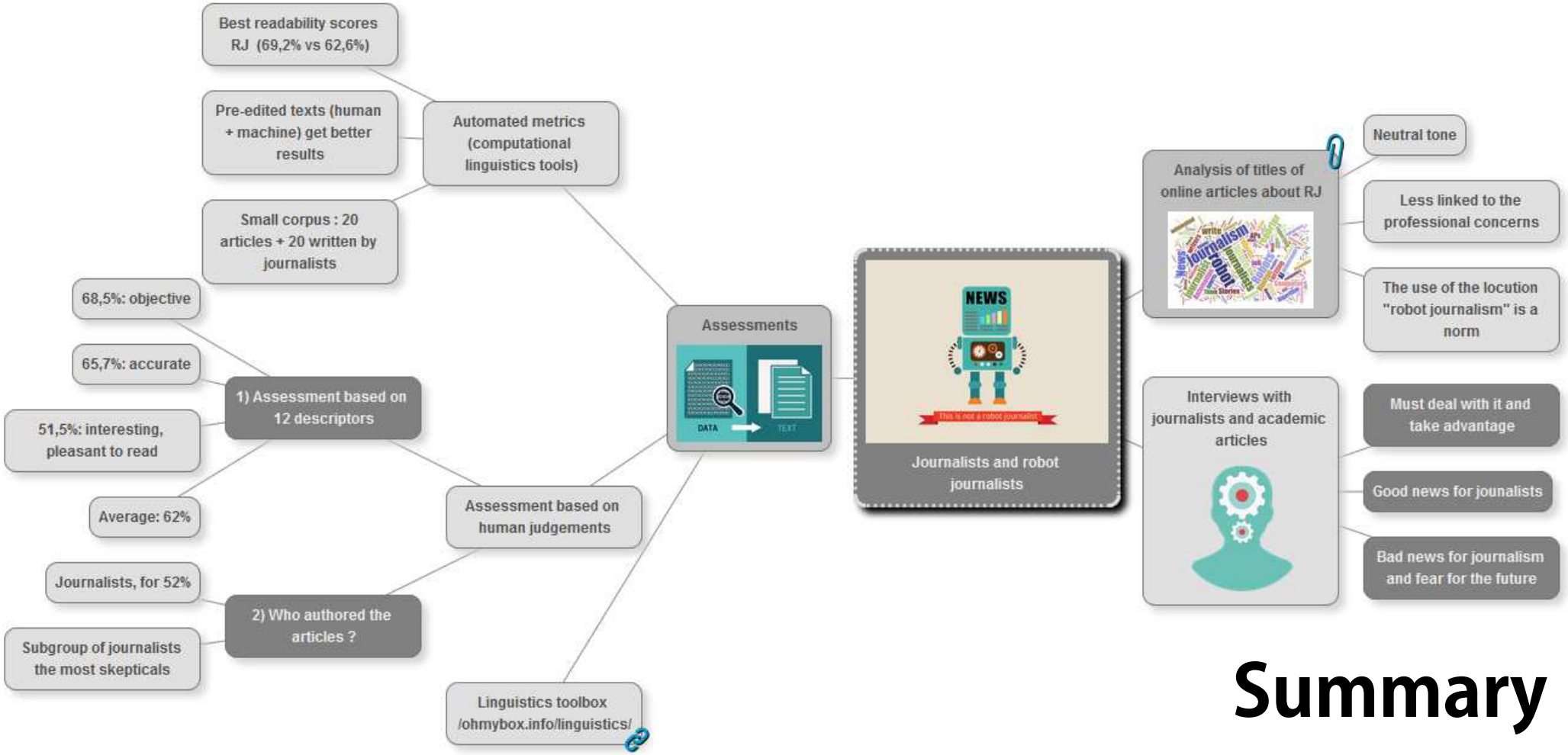


52% written by a human on 3 articles

76% written by a human on 2 articles (best scores)



Perception of automated news articles



Summary

Perception of automated news articles

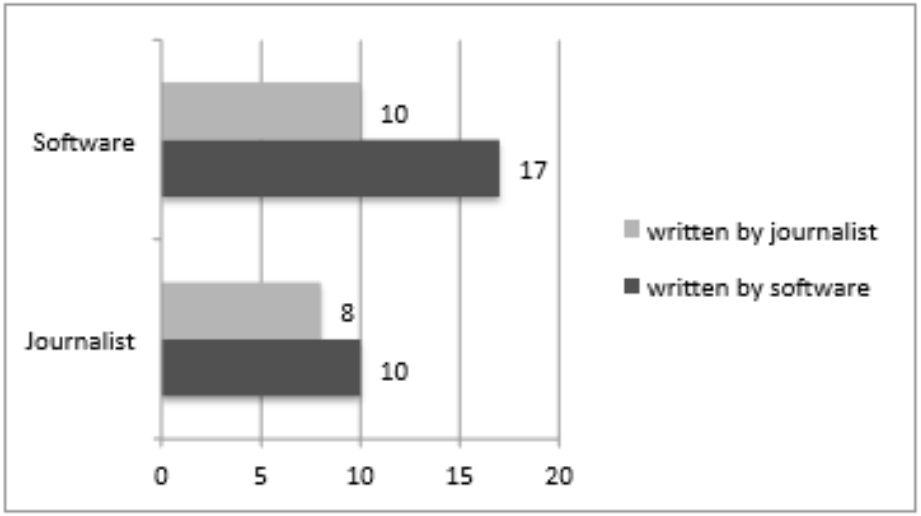
Enter the robot journalist (Clerwall, 2014, Sweden)

46 students participated in the test, 30 women (65 percent) and 16 men (35 percent).

2 texts submitted: written by a software and written by a journalist

The **text written by a journalist** scores higher on coherence**, "well written", "clear",** and on being **pleasant to read**.

The **software-generated text** scores higher on other descriptors, such as being **descriptive** (whether or this is a positive may of course be a matter of personal preferences), **informative**, **trustworthy**, and **objective** (credibility).



Journalist versus news consumer: The perceived credibility of machine written news (Van der Kaa, Krahmer, 2014, Netherland)

Audience study with a focus on the credibility of automated journalism

232 native Dutch speakers (the language of the experiment) took part in this research, and among them were 64 journalists

Within the **group of news consumers**, no main effect was found. News consumers perceive the trustworthiness and expertise of the computer writer and journalist equally. In general, they were neutral about the levels of expertise of both the computer writer and journalist writer. No differences in the perceptions of news consumers regarding the credibility of machine-written news articles

Perception of automated news articles

Journalist versus news consumer: The perceived credibility of machine written news (Van der Kaa, Krahmer, 2014, Netherland)

Within the **group of journalists**, there was no effect on the perceived expertise of the news source. In general, journalists were slightly positive about the levels of expertise of the computer writer and the journalist writer.

RKC Waalwijk versus PEC Zwolle: match ends in a draw

RKC Waalwijk visited PEC Zwolle and drew. The duel ended in one – all. Twelve thousand spectators came to the Ijsseldelta stadium.

The team from Zwolle took the lead after 44 minutes with a goal by Saymak. Four minutes later, Joachim from RKC Waalwijk equaled the score.

The match was officiated by referee Kamphuis. He did not issue any red cards. Tomas and De Boer of PEC Zwolle picked up a yellow card.

This article was written by a computer.

Perception of Automated Computer-Generated News: Credibility, Expertise, and Readability (Haim, Graefe, Haarmann, Brosius, 2015)

986 subjects rated two articles on credibility, readability, and journalistic expertise

Computer-written news tends to be rated higher than human-written news in terms of credibility

News consumers get more pleasure out of reading human-written as opposed to computer-written content. Articles are consistently perceived more favorably if they are declared as written by a human journalist

Differences in terms of perceived credibility and expertise tend to be small. A possible explanation for the small differences is that algorithms strictly follow standard conventions of news writing and, as a result, computer-written stories reflect these conventions.

When Reporters Get Hands-on with Robo-Writing (Dörr, Thurman, Kunert, 2017)

Analyze professional journalists' experiences with, and opinions about, the technology.

Participants were drawn from a range of news organizations—including the BBC, CNN, and Thomson Reuters—and had first-hand experience working with robo-writing software provided by one of the leading technology suppliers.

Results reveal journalists' judgements on the limitations of automation, including the nature of its sources and the sensitivity of its "nose for news"

Journalists believe that automated journalism will become more common, increasing the depth, breadth, specificity, and immediacy of information available

Perception of automated news articles

When Reporters Get Hands-on with Robo-Writing (Dörr, Thurman, Kunert, 2017)

Journalists' reactions were largely negative but many writers did appreciate the potential positives of automated journalism: could "present the facts as they are" without "manipulation" ; could help to quickly break stories initially, before real journalists took the helm for further coverage

Perception of automated news articles

To conclude, few recommendations

1. NLG softwares should be considered in all cases as a support for journalism.

2. NLG softwares should not be a cost-cutting way that would be to the detriment of journalistic employment

3. For the sake of transparency, the data used must be traceable (reference to data producer).

4. Data sources must be accurate, reliable and up-to-date. Fact-checking procedures should be implemented in particular when data are coming from third-parties knowledge bases.

5. Any human being (programmers, linguists…) involved in an NLG process must take into account the ethical dimensions governing journalism.

To conclude, few recommendations

6. Journalists must be involved in this process to remain active actors

7. The structure of narratives must always be adapted to the types of data processed and to their application domain

8. Automated contents should not used to much repetitive structures and offer a certain variety in the narratives proposed

9. Implementation of NLG software should be tested and evaluated with the concerned audiences

10. Automated contents should be always be mentioned as written by a software

Thank you!