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.
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.
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)

When computational linguistics meets journalism
Theoretical process

Pipeline architecture (Dale and Reiter)

When computational linguistics meets journalism
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

When computational linguistics meets journalism
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

When computational linguistics meets journalism
NLG and journalism: meeting point

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

When computational linguistics meets journalism
A brief history of NLG

- Field of natural language processing (NLP)
- Growing since the late 1960’s: playing field for researchers
- 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...

When computational linguistics meets journalism
NLG and journalism: a growing history

When computational linguistics meets journalism

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).
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.
NLG and journalism: a growing history

When computational linguistics meets journalism

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.
NLG and journalism: a growing history

When computational linguistics meets journalism

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.”
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

A shallow magnitude 3.4 earthquake was reported Tuesday evening 5 miles from Niland, Calif., according to the U.S. Geological Survey. The temblor occurred at 10:00 p.m. PST at a depth of 0 miles.

According to the USGS, the epicenter was 14 miles from Brawley, Calif., 22 miles from Imperial, Calif. and 26 miles from El Centro, Calif.

In the past 10 days, there has been one earthquake of magnitude 3.0 or greater centered nearby.

This information comes from the USGS Earthquake Notification Service and this post was created by an algorithm written by the author.
NLG and journalism: a growing history

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 processing”, 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

SEPTEMBER 2014
Sport Informations Dienst (SID), a German subsidiary of Agence France Presse (AFP), announces the automated production of sports reports in thirteen languages. It's a pilot project achieved with the technology of the German company Aexea.

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

When computational linguistics meets journalism

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. 50,000 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.
NLG and journalism: a growing history

When computational linguistics meets journalism
NLG and journalism: a growing history

MAY 2015
Publication of "The robotic reporter" by Matt Carslon (Nov.2014) in Digital Journalism. The author defines automated journalism as "an algorithmic processes that convert data into narrative news texts with limited to no human intervention beyond the initial programming choices".
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 the media organisation NDP (members: Persgroep, RTL, ANP). Some research will be carried out at the Telegraaf (newspaper).
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.
NLG and journalism: a growing history

When computational linguistics meets journalism
NLG and journalism: a growing history

OCTOBER 2015
Yandex, the Russian group well known for its search engine, announces the launch in November of a new automated news service, which the company calls "the future of news agency". The areas covered are the weather forecasts and traffic. Yandex will also propose this service to Interfax (news agency), which starts to automate contents through its subsidiary Finmarket.

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

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

November 2015

The news agency Xinhua has announced the adoption of natural language generation for sports and finance reports. The software, called "Kuaibi Xiaoxi" ("Little Xinhua Who Writes Fast"), works in Chinese and English.
The regional elections in France give rise to a new partnership between Syllabs and Le Monde. The French company has also signed for the occasion, a partnership with France Bleu, L'Express and Le Parisien.
NLG and journalism: a growing history

When computational linguistics meets journalism
NLG and journalism: a growing history

When computational linguistics meets journalism

JANUARY 2015
NLG softwares are continuing their breakthrough in the German media. Since 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.
NLG and journalism: a growing history

JANUARY 2016
A tribute to Marvin Minsky, a pioneer in the field of artificial intelligence, was given by Wordsmith (Automated Insights) with the writing of an obituary for Wired. The result was considered as conclusive.
NLG and journalism: a growing history

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.

Image: CanStockPhoto
NLG and journalism: a growing history

When computational linguistics meets journalism

APRIL 2016

"Digital News Initiative", a program launched by Google to support innovative journalism, have given a grant of € 46,200 to Journalism++ in Sweden for the development of MARPLE. Its principle: the use of statistics and public data to produce information. A thousand articles are planned to be written within three years.
NLG and journalism: a growing history

Bloomberg NEWS

APRIL 2016
Bloomberg news agency launches a 10-person team to study how to increase the automation of writing and reporting. Chief-editor John Micklethwait said that the employment of the 2,400 journalists and analysts is not threatened and that “done properly, automated journalism has the potential to make all our jobs more interesting.” The time spent laboriously trying to chase down facts can be spent trying to explain them. We can impose order, transparency and rigor in a field which is something of a wild west at the moment. Bloomberg already use automation for news alerts, customized news and trending stories.

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 journalism. 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

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

When computational linguistics meets journalism
NLG and journalism: a growing history

When computational linguistics meets journalism

DECEMBER 2016
The first experiment of Finnish robotic journalism is launched by Yle (Yle Urheilu) with Voitto. Its purpose is to automate the reporting of sports events based on a statistical treatment (relating to the meeting but also those of previous seasons). Each report is published at the blown the final whistle of the match.
NLG and journalism: a growing history

JANUARY 2017
The adventure continues in China where the Southern Metropolis Daily published its first report written by a software. The story counted 500 mots, and "summarizes what train tickets are most in demand over the Lunar New Year holiday".
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 launched 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 Israeli 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...

When computational linguistics meets journalism
...and there are also (almost) „real“ robot journalists

First android newscaster in Japan (2014)

When computational linguistics meets journalism
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 to the 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

When computational linguistics meets journalism
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

When computational linguistics meets journalism
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.

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
When computational linguistics meets journalism

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

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>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Creativity allows human journalists to go beyond cliché and add humour to their stories</td>
<td></td>
</tr>
<tr>
<td>- Flexibility allows human journalists to cover non-routine stories like breaking news</td>
<td></td>
</tr>
<tr>
<td>- Analytical skills allow human journalists to go beyond description and to provide in-depth coverage</td>
<td></td>
</tr>
<tr>
<td>- Human journalists have higher marginal costs than automated content creation</td>
<td></td>
</tr>
<tr>
<td>- Human journalists cannot provide the same breadth of coverage as automated content creation</td>
<td></td>
</tr>
<tr>
<td>- Human journalists cannot compete on speed with automated content creation</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Opportunities</th>
<th>Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Automating routine stories gives journalists more time for research and in-depth reporting</td>
<td></td>
</tr>
<tr>
<td>- Competition from automated content creation pushes human journalists to do a better job</td>
<td></td>
</tr>
<tr>
<td>- Automated content creation can help to cover stories for small audiences which now go uncovered</td>
<td></td>
</tr>
<tr>
<td>- Automated content creation may put journalists doing routine tasks out of work</td>
<td></td>
</tr>
<tr>
<td>- Automated content creation can be applied beyond sports reporting and also challenge the jobs of journalists in finance or real estate</td>
<td></td>
</tr>
<tr>
<td>- Automated content creation raises new ethical questions, including issues like transparency and copyright</td>
<td></td>
</tr>
</tbody>
</table>

Source: based on 50 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.

When computational linguistics meets journalism
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)

Tone of the title

- Neutral: 64%
- Positive: 18%
- Négative: 18%

Few related to the identified issues (%)

- Job: 8.14%
- Professional identity: 9.3%
- Change/future: 14.06%
- Man/machine: 18.63%

Perception of automated news articles
What do journalists think about RJ?

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

- 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)

respondents = experts in writing (media field)
What do journalists think about RJ?

<table>
<thead>
<tr>
<th>Quality Attribute</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliable</td>
<td>64.66%</td>
</tr>
<tr>
<td>Trustworthy</td>
<td>61.63%</td>
</tr>
<tr>
<td>Objective</td>
<td>68.46%</td>
</tr>
<tr>
<td>Boring</td>
<td>64.46%</td>
</tr>
<tr>
<td>Interesting</td>
<td>51.51%</td>
</tr>
<tr>
<td>Pleasant to read</td>
<td>51.52%</td>
</tr>
<tr>
<td>Clear</td>
<td>62.15%</td>
</tr>
<tr>
<td>Informative</td>
<td>63.52%</td>
</tr>
<tr>
<td>Well-written</td>
<td>60.39%</td>
</tr>
<tr>
<td>Complete</td>
<td>65.17%</td>
</tr>
<tr>
<td>Accurate</td>
<td>65.69%</td>
</tr>
<tr>
<td>Coherent</td>
<td>64.55%</td>
</tr>
</tbody>
</table>

Articles about economics

Judges did not know the real subject of the experience (assessment of the quality of articles on economics published online)
 Were they able to recognize the software?

Who authored the news articles?

<table>
<thead>
<tr>
<th>Sample</th>
<th>Software</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>48.0%</td>
<td>52.0%</td>
</tr>
<tr>
<td>Sample 2</td>
<td>22.7%</td>
<td>77.3%</td>
</tr>
<tr>
<td>Sample 1</td>
<td>25.3%</td>
<td>74.7%</td>
</tr>
</tbody>
</table>

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

Best readability scores
RJ (69.2% vs 62.6%)

Pre-edited texts (human + machine) get better results

Small corpus: 20 articles + 20 written by journalists

68.5% objective
65.7% accurate
51.5% interesting, pleasant to read
Average: 82%

Journalists, for 52%
Subgroup of journalists the most skepticals

Linguistics toolbox
johnnymbox.info/linguistics/

1) Assessment based on 12 descriptors
2) Who authored the articles?

Assessment based on human judgements

Automated metrics (computational linguistics tools)

Analysis of titles of online articles about RJ

Neutral tone
Less linked to the professional concerns
The use of the location "robot journalism" is a norm

Interviews with journalists and academic articles

Must deal with it and take advantage
Good news for journalists
Bad news for journalism and fear for the future
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
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 equalled 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.

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
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.

**ROBO-JOURNALISM’S TOP LIMITATIONS ACCORDING TO JOURNALISTS WHO’VE WORKED WITH THE TECHNOLOGY:**

1. Its reliance on single, isolated data streams.
2. The one-dimensional nature of the quantitative data feeds it relies on.
3. The difficulties of interrogating that data.
4. The lack of human angles in the texts generated.
5. The requirement to template stories—predicting “top lines”—in advance.
6. The difficulty of working creatively with the data in the templating process.
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 use too 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!

www.ohmybox.info - @ohmyshambles