Addressing Micro-Audiences at Scale
How Tamedia generated about 40,000 articles in five minutes to report on Swiss popular vote results at the municipalities level

Titus Plattner†
Digital editorial development
Tamedia
Lausanne, Switzerland
titus.plattner@tamedia.ch

Didier Orel
Data analytics team
Tamedia
Zürich, Switzerland
didier.orel@tamedia.ch

ABSTRACT
Should the Swiss cows keep their horns? Or to be precise: must the farmers be financially encouraged not to de-horn the cattle and goats? On November 25th 2018, among two other proposals, this was the crucial question Swiss voters had to decide on. The initiative committee believed that it is inhumane to remove their horns. Gentrified cities and remote villages in the mountains predominantly supported the horn bonus, but the suburban regions said no. Finally, 55 percent of the voters rejected this idea. And on that same day, “Tobi”, Tamedia’s new textbot, generated for the first time automated articles detailing the vote results for each of the 2222 municipalities of the country. This experiment allowed us to address a long tail audience, and improve the user engagement thanks to personalization.

CCS CONCEPTS
• Applied computing → Media arts; • Information systems → Information systems applications • Social and professional topics → Automation; automated journalism, news values, decision tree algorithm

KEYWORDS
Automated journalism, textbot, poll results, audience engagement, long tail audience, user feedback

ACM Reference format:
Titus Plattner and Didier Orel. 2018. Addressing micro-audiences at scale. How Tamedia generated about 40,000 articles in five minutes to report on Swiss popular vote results at the municipalities level. In (Computation+Journalism ’19). Miami, FL, USA.

1 Scaling
Tamedia is the biggest private media group in Switzerland, which owns more than 30 newspapers in the country. After two low profile experiments in the cantons of Bern and Zurich in June and September 2018, our textbot was used on a national level for the first time on November 25th. The users had the possibility to indicate beforehand their vote preferences, to get an even more personalized text, describing if they were or were not on the side of the majority for each of the three proposals submitted to the vote that day. The articles were written both in German and French.

In total, the decision-tree algorithm we nicknamed “Tobi” generated 39,996 different articles (2222 “communes” in the Swiss terminology x (23 + 1) vote options x 2 languages). The articles gave the local results compared to the cantonal and national results; showed if the commune was one of the most or least favorable to an object of the vote; gave some additional localized information explaining what could have influenced the result; “commented” or “analyzed” to some point the outcome; and gave the details of the local voter turnout. On average, the articles were about 250 words long (2 minutes read), accessible on most of Tamedia’s online platforms, from the national daily newspapers like Tages-Anzeiger¹ or Tribune de Genève² to the local ones, like Zürcher Unterländer. See Figure 1 for an example of one of the articles in German.

1 https://interaktiv.tagesanzeiger.ch/2018/tobi/
2 https://interaktiv.idg.ch/2018/tobi-nov/
Behind the bot, experienced reporters

It would have been nonsensical to produce forty thousand articles manually, by reporters. A human, spending about 10 minutes per article, would need more than 800 office days. “Tobi” generated them in less than 5 minutes (albeit with a fair bit of design and development time before that).

To prepare the templates, we used a commercial solution from Automated Insights and we built a specific back-end and front-end for this project from scratch. The way to deliver these articles is as important as the articles themselves. As more than 80% of our audience is on mobile devices, we focused on the best possible user experience on smartphones.

A small team of five experienced political reporters spent about 2-3 days each to imagine the different scenarios, conceive the decision tree and pre-write text bricks, with the aim to have an output that seems as “human” as possible. We tried to avoid repetition and wanted our texts to be more than just an enumeration of percentages, ranks, and a number of voters. To achieve this, we added transition sentences to let the reader breath, introduced analysis where it was possible and even inserted some glittering moments in the articles, for example with a well-chosen word or even some humor. To accomplish this you need good journalists and to give them time. Some of these text bricks can be used or adapted for future voting events, which will reduce the human effort needed.

In addition to that work on the texts, a data journalist processed the official vote data and enhanced them with some context related statistics such as previous vote results, the number of farmers versus workers or white collars, urbanisation rates, etc. The different grammatical construction between French and German, as well as the cultural differences, pushed us to tailor the templates by language. To accomplish it, we also had to adapt the database linked to the template. In French, for example, we had to keep in a separate column the gender and number of every municipality name, which can vary in certain circumstances. In another column, we kept the demonyms, the word that identifies residents of a particular place (“les Genevoises et les Genevois” live in Geneva would be an easy one; “les Agaunoises et les Agaunois” live in Saint-Maurice would be a more tricky situation, that no algorithm could predict). In the end, the table contained more than 300,000 cells (137 variables per commune).

3 Long tail audience

In total, more than 100,000 users read at least one of the automated articles between Sunday at 5 pm when the results were published and the following Tuesday morning. In the Swiss context, this is a huge amount. The 55 communes with the biggest audience gathered together 25% of all requests. The 200 following communes gathered another 25% of the audience and the remaining 50% of the audience was distributed on 1935 communes. Among the 2222 communes that Switzerland counts, only 32 had zero requests, what shows how widespread the reach was (See Figure 2).
Addressing Micro-Audiences at Scale

With this experience on vote results, Tamedia’s goal is to deliver specific public-interest content to our users in municipalities who ordinarily get only poor media coverage. For example, Maschwanden in the canton of Zurich, 620 residents, got its last substantial article in the national AND local press over 30 months ago, in July 2016. On November 25th, 16 users requested the algorithmically generated article about Maschwanden. The penetration rate in this small municipality was about 2.5 times higher than on average. We also noted good penetration rates in regions where Tamedia’s platforms are less present.

Of course, automated articles will never provide the same information as a handcrafted, well-researched article. But we hope that these kind of automated articles will, in the long run, help us to reinforce the user engagement by reporting more often on what’s very close to users.

4 Measuring the user feedback

In the user interface, we also have implemented a tool to let our users give some feedback on a five star scale. About 2% of the users decided to give a rating, which is in the norm of what we experienced on our platforms with other projects.

Compared to other projects, we consider the feedback ratings of, 4.5, 4.1 and 3.8 respectively (See Figure 4), to be very good. Especially the one about the text quality (4.1), where our expectations for an automated article were much lower. In comparison to the experiment we made in September, we improved the usefulness rate by 0.3 points and the two others by 0.2 points, what is very encouraging. There is still a long way to go but we think that, in some fields, it is possible to produce automated articles that you can barely differentiate from a reporter-written report.

Tamedia is also planning automated texts for sports results or quarterly earnings reports, for example. Other media organizations, with Associated Press at the vanguard since 2014, already use this approach. But we think that a personalized distribution of these texts – with for example personalized push alerts or a tailored feed – will allow leveraging this long-tail audience.

5 Transparency and humility

Transparency on how those texts are produced is key. In an era where both algorithmic black boxes and journalists are more and more scrutinized and challenged, clarity on the whole process will help to keep the trust of the users. A certain humility is also required. On all Tamedia’s articles, “Tobi” had a feedback button, asking users to report any errors to us. On the polling day, an value inversion in one sub-branch of the template in German could be fixed within a few minutes after users reported the error.

During the last experiment, the byline of the articles was “Tobi, your textbot”: this had the advantage to show very clearly that this was an automated article. We even personified Tobi by creating a mascot. The downside of that was that the journalists who worked a lot to write the template, only appeared in an Impressum, at the very bottom of the page. Because we believe that the most value comes from the quality of the templating, achieved by journalist, we may reconsider who is in the byline in the future.

ACKNOWLEDGMENTS
Roland Bapst, Daniel Barben, Mathias Born, Renaud Bournoud, Marc Brupbacher, Yvan da Silva, Beni Gafner, Simon Koch, Kaspar Manz, Lucie Monnat, Rita Frommenwiler, Viviane Futterknecht, Fabian Renz, Mathieu Signorell and Philippe Sordet, all, big or small, contributed to this project.

Figure 3: Penetration rates of Tobi’s article requests

Figure 4: The user feedback scale for “Tobi”