

MARIUSZ HERMA

Tygodnik „Polityka”

dziennikarz i założyciel globalnego serwisu beehype

## TŁUMACZENIA AUTOMATYCZNE: JAKOŚĆ I ICH WPŁYW NA KOMUNIKACJĘ W SIECI

Prototyp stosowanych obecnie tłumaczeń automatycznych zaprezentowali już jesienią 2012 roku programiści firmy Microsoft, odpowiadającej za serwis Bing Translate [Rashid 2012]. Obecnie za najbardziej zaawansowaną uznaje się jednak konkurencyjną usługę Google Translate. Obsługuje ona sto trzy języki<sup>1</sup>, którymi posługuje się aż 99% internautów. Liczba języków obsługiwanych przez Google Translator jest prawie dwukrotnie większa niż liczba języków Bing Translate. Osoba korzystająca z Translate może wypowiedzieć kilka zdań do mikrofonu komputera lub smartfonu, a jego wypowiedź niemal natychmiast zostanie zamieniona na tekst, przetłumaczona na wybrany język i „odczytana” przez automatycznego lektora. Osoby posługujące się różnymi językami mogą zatem usiąść przy jednym stole i rozmawiać ze sobą, mimo że każda pozostanie przy własnym języku. Każdy z nas może dysponuje więc niejako osobistym tłumaczem symultanicznym – tyle że znającym ponad sto języków.

### LUDZIE I TŁUMACZE

Historię tłumaczeń automatycznych zazwyczaj dzieli się na trzy okresy. Za pierwszą generację uznaje się tłumaczenia z okresu zimnej wojny, gdy szczególnie Amerykanie zainteresowani byli szybkim wglądem w dokumenty rosyjskie czy chińskie. Wymagało to jednak załadowania do komputera słownika danego języka oraz dokładnych zasad jego gramatyki. Przetłumaczony tekst wymagał następnie interpretacji lingwistów.

Na przełomie XX i XXI wieku pojawiła się druga generacja tłumaczeń automatycznych, a za ich głównego reprezentanta uznaje się rozwijany do 2006 roku Google Translate. Systemy tłumaczeniowe drugiej generacji bazowały na statystyce: porównywały wykonane ludzką ręką przekłady tych samych tekstów, na przykład dokumentów Organizacji Narodów Zjednoczonych (tłumaczone na co najmniej sześć oficjalnych języków ONZ). Szczególnie przydatne okazały się tutaj światowe bestsellery literackie, takie jak chociażby seria „Harry Potter”, którą przetłumaczono na siedemdziesiąt

---

<sup>1</sup> Sto cztery, jeśli oddzielnie liczyć chiński tradycyjny i chiński uproszczony.

języków, a w wydaniu polskim liczy ponad cztery tysiące stron. Równoległe z pracami nad usługą Translate firma Google rozwijała projekt Books. Jego celem było „zeskanowanie wszystkich książek świata” – do 2015 roku udało się to z ponad dwudziestoma pięcioma milionami tytułów w rozmaitych językach. Pozyskane w ten sposób dane niejako przy okazji wspierały system Translate.

Zaletą drugiej generacji tłumaczeń było to, że nie wymagały wprowadzenia do systemu słownika oraz gramatyki danego języka. Komputer niejako „bezmądrze” – bazując wyłącznie na statystyce – naśladował ludzi, tłumaczy. Jednocześnie rozwiązaniem to miało dwie zasadnicze wady.

Po pierwsze, jakość tłumaczeń zależała od liczby i jakości dokumentów dostępnych w danym języku. O ile nie stanowiło to problemu na przykład dla angielskiego, francuskiego czy niemieckiego, o tyle stosunkowo niewiele dokumentów i książek tłumaczy się na język arabski, mimo że posługuje się nim nawet czterysta milionów osób. O jeszcze mniejszej dostępności można mówić w przypadku języków lokalnych, takich jak na przykład fryzyjski, cebuański czy sotho. Jako anegdotę można przytoczyć historię włączenia języka kazachskiego do bazy Translate. Programiści Google poinformowali, że jest to warunkowane zgromadzeniem dwóch milionów par zdań (po kazachsku i angielsku). Sam przywódca kraju wygłosił więc orędzie do rodaków, w którym zaapelował, by przysyłali do Google takie materiały. Firma otrzymywała 100 tys. par zdań dziennie i wskutek tego w 2014 roku kazachski dołączył do listy języków obsługiwanych przez Google Translate.

Drugą wadą systemów tej generacji było to, że oferowały tłumaczenia bez uwzględniania kontekstu wypowiedzi, a dłuższe frazy musiały dzielić na fragmenty po pięć słów. W efekcie „Google Translate mogło w dwie sekundy przetłumaczyć akapit, który mu zadano, jednak później potrzeba było dwóch minut, aby rozszyfrować sens otrzymanego przekładu” – przyznawał w rozmowie ze mną Barak Turovsky, dyrektor tego projektu. Tłumaczenia tej generacji wyłącznie na podstawie statystyk decydowały, jak należy przetłumaczyć na przykład słowo *zamek* albo wypowiedziane na głos *może/morze*.

Z tego względu w drugiej dekadzie XXI wieku rozpoczęto prace nad tłumaczeniami trzeciej generacji. Chodzi o tak zwane *neural machine translation* – tłumaczenia maszynowe neuronowe. Miały one brać pod uwagę kontekst wypowiedzi i patrzeć na nią jako na całość: na całe zdanie, akapit, a najlepiej na cały dokument. Google Translate zaczęło wdrażać tę technologię w 2016 roku. Dla języka polskiego był to marzec 2017 roku, a od 2018 roku tłumaczono tak wszystkie dostępne języki. Powołując się na własne testy, eksperci Translate twierdzili, że „przełączenie tłumaczeń na sztuczną inteligencję” pozwoliło ograniczyć liczbę błędów o 60% [Wu i in. 2016]. Oznaczało większy postęp niż przez poprzednie dziesięć lat.

W przypadku języka polskiego badani użytkownicy uznawali za „bardzo dobre” 72% tłumaczeń z polskiego na angielski (wcześniej było to 55%), a z angielskiego na polski – 48% (wcześniej 18%). Oczywiście w dłuższych zdaniach, szczególnie tych o zabarwieniu literackim, Translate wciąż popełnia błędy. Ale na przykład tłumaczenie artykułów z anglojęzycznej prasy często nie różniłyby się istotnie od tych wykonanych przez człowieka. Najwyższą jakość tłumaczeń można uzyskać tylko dla języka angielskiego, ale przedstawiciele Google zauważają, że ok. 50% treści w Internecie publikuje się właśnie w tym języku, choć posługuje się nim zaledwie co piąty internauta.

**BBC**  
**NEWS**

**Olga Tokarczuk and Peter Handke win Nobel Prize for Literature for 2018 and 2019**



Olga Tokarczuk won for 2018 and Peter Handke is 2019's winner

Polish author Olga Tokarczuk and Austria's Peter Handke have been awarded the Nobel Prize for Literature.

Two winners were named - one for 2019 and one for 2018 - because the prize was not awarded last year.

The Swedish Academy, which oversees the prestigious award, suspended it in 2018 after a sexual assault scandal.

Tokarczuk, who also **won the Man Booker International Prize last year**, was awarded the 2018 Nobel Prize, with this year's Nobel going to Handke.

**BBC**  
**NEWS**

**Olga Tokarczuk i Peter Handke zdobywają Nagrodę Nobla w dziedzinie literatury na rok 2018 i 2019**



Olga Tokarczuk won for 2018 and Peter Handke is 2019's winner

Polska autorka Olga Tokarczuk i austriacki Peter Handke otrzymali Nagrodę Nobla w dziedzinie literatury.

Wyłoniono dwóch zwycięzców - jednego na 2019 r. i jednego na 2018 r. - ponieważ nagrody nie przyznano w ubiegłym roku.

Szwedzka Akademia, która nadzoruje prestiżową nagrodę, zawiesiła ją w 2018 r. Po skandalu z napaścią seksualną.

Tokarczuk, który w ubiegłym roku zdobył także **nagrodę Man Booker International**, otrzymał Nagrodę Nobla 2018, a tegoroczny Nobel trafił do Handke.

Rysunki 1, 2

Oryginalna wersja artykułu BBC o nagrodzie Nobla dla Olgi Tokarczuk i Petera Handkego (z dnia 10 października 2019 r.) oraz automatyczne tłumaczenie tego artykułu wykonane przez Google Translate

## TŁUMACZENIA WYSOKIE I NISKIE

W ostatnich latach najliczniejszą grupą użytkowników Translate byli Brazylijczycy, relatywnie chętnie korzystają z niej także mieszkańcy Indii czy Indonezji. Szukając przyczyn tego stanu rzeczy, eksperci Translate zauważyli, że w krajach rozwijających się znajomość języka angielskiego świadczy o statusie społecznym. Kompetencje językowe najczęściej zaś się demonstruje w mediach społecznościowych. Translate i inne tego typu serwisy znakomicie nadają się do tłumaczenia krótkich fraz (powitań i pożegnań, gratulacji i wyrazów zdziwienia), które są następnie publikowane na Facebooku, Twitterze czy Instagramie. Jednocześnie automatyczne tłumaczenia pozwalają zrozumieć, co (również po angielsku) napisali lub odpisali nasi znajomi.

Dość szybko zauważyły to główne serwisy społecznościowe i same wprowadziły możliwość automatycznego tłumaczenia postów – pod wpisami w innym języku widzimy dziś przycisk „przetłumacz”, można też zażyczyć sobie automatycznego tłumaczenia takich treści. Facebook początkowo korzystał z usługi Microsoft Bing, ale od ponad trzech lat rozwija własny system. Codziennie odpowiada on na cztery do pięciu miliardów zapytań (nie licząc tłumaczeń na Instagramie, który także należy do Facebooka). Z algorytmów Bing wciąż korzysta zaś Twitter.

Tłumaczenia automatyczne umożliwiają także wgląd w „życie” innych krajów. Jednym kliknięciem można sprawdzić, co napisali nasi znajomi z Erasmusa, jak skomentowała swoje nowe zdjęcie nasza ulubiona artystka z Brazylii – i o czym śpiewa



Rysunki 3, 4

Przykład wykorzystania aplikacji do tłumaczenia automatycznego na żywo: napis filmowany lub fotografowany za pomocą komórki na bieżąco zastępowany jest tłumaczeniem na wybrany język

w nowej piosence, łatwo również sprawdzić, czy najpopularniejszy obecnie hasztag na Twitterze – pisany po koreańsku, turecku czy arabsku – odnosi się do nowego odcinka popularnego serialu, czy też chodzi o nowe protesty przeciwko władzy.

Przeźren wirtualną tłumaczenia automatyczne w dużej mierze już więc zagarnęły. Ich ciągły rozwój rodzi jednak pytania o dalszą ekspansję. Sami eksperci firm Google czy Microsoft podkreślają, że nie mają ambicji – ani możliwości – by doprowadzać do sytuacji, w której zawód tłumacza straci sens. Zwracają uwagę na podział tłumaczeń na „niskie” oraz „wysokie”. Te drugie, dotyczące na przykład literatury pięknej czy negocjacji biznesowych, zawsze pozostaną według nich specjalnością ludzi. W takich okolicznościach krytyczne są bowiem rozmaite niuanse, etykieta, kontekst itd. Z kolei tłumaczenia „niskie” w praktyce już zostały zdominowane przez algorytmy. Chcąc przetłumaczyć przepis kuchenny albo instrukcję obsługi, internauci odruchowo otworzą serwis Translate lub Bing – a jeszcze kilkanaście lat temu sięgnęliby po słownik lub zadzwonili do znajomego, który włada danym językiem.

Rodzi się pytanie o przestrzeń pomiędzy tymi dwoma rodzajami tłumaczeń. Jaką jej częścią zawładną algorytmy? Wiele osób jest zaskoczonych chociażby tym, jak szybko tłumaczenia automatyczne wyparły rozmówki turystyczne, szczególnie odkąd Translate zaczęło tłumaczyć nie tylko słowa, które wpisujemy w okienko – można też posłużyć się głosem lub sfotografować zawartość menu. Oryginalny tekst zostanie zastąpiony (na samym obrazie) słowami w naszym ojczystym języku. Nie trzeba za to płacić, a w przypadku większości języków niepotrzebny jest także dostęp do Internetu: potrzebne dane można pobrać do pamięci urządzenia przed wyjazdem.

## **BABEL 2.0**

Po raz pierwszy w historii ludzkości znajomość języków obcych przestaje zatem być konieczna do tego, by porozumieć się z obcokrajowcami. Równie łatwy staje się dostęp do wszelkiego rodzaju treści – tekstów, filmów, piosenek – w innych językach. Również tych uznawanych za niszowe, które zgłębiają tylko wyspecjalizowani tłumacze i pasjonaci.

W jaki sposób tak duża dostępność tłumaczeń automatycznych, przy wciąż rosnącej ich jakości, wpłynie na całą cywilizację? Ich ekspansja dokonuje się niejako w cieniu innych nowinek technologicznych, takich jak media społecznościowe. Z perspektywy stuleci raczej niż lat czy dekad tak głęboka zmiana może się okazać znacznie bardziej doniosła niż wynalazek mediów społecznościowych, nawigacji GPS czy streamingu filmów i muzyki, choć na razie skutecznie odwracają one naszą uwagę od tej cichej językowej rewolucji.

## **ŹRÓDŁA**

- Rashid R., 2012, Speech Recognition Breakthrough for the Spoken, Translated Word, <https://www.youtube.com/watch?v=Nu-nlQqFCKg> (dostęp: 25.10.2019).
- Wu Y. i in., 2016, Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, <https://ai.google/research/pubs/pub45610> (dostęp: 25.10.2019).



## DYSKUSJA DO REFERATÓW

**Kazimierz Sikora:** Do wystąpienia pana Troszyńskiego. Myślę, że warto przywołać parę oczywistych prawd. Przeklinanie w miejscach publicznych na przykład jest karalne, a mandat może wynieść w różnych okolicznościach do pięciuset złotych. A tymczasem na przykład na stadionie półtora tysiąca osób używa niewybrednych wulgaryzmów. Ta sytuacja jest w największym stopniu demoralizująca, bo rodzi się przekonanie, że coś takiego jak przestrzeń publiczna nie istnieje. Wydaje mi się, że przestrzeń Internetu można traktować analogicznie jak przestrzeń publiczną. Jeśli tak jest, to wystarczy – cytując klasyków – po prostu przestrzegać prawa. A więc wytropienie hejtera wcale nie jest takie trudne.

**Maciej Piasecki:** Hejting staje się już nie tylko rozrywką, ale też dobrze płatnym zawodem (niektórzy zarabiają po kilkaset tysięcy miesięcznie); zarabia się, organizując zrzutkę internetową, czyli tak zwany *crowdfunding*. Jak w każdym zawodzie, pojawia się tu specjalizacja. Na przykład są osoby specjalizujące się w „podpalaniu” (*flamer* – ‘podpalacz’), które nie sieją nienawiści, tylko napuszczają ludzi na siebie. Innym „specjalistą” jest patostreamer – „patologiczny streamer”, czyli osoba, która w celach patologicznych zamieszcza filmiki w Internecie, twierdząc, że to jest na żywo (choć niekoniecznie jest to relacja na żywo), na przykład namawia nastolatkę do tego, by się rozebrała przed kamerą, a potem wstawia taki film do Sieci. Czy istnieje jakaś analiza takich specjalności?

**Renata Przybylska:** Chciałam zapytać pana profesora Filiciaka, czy przewiduje pan powstanie takiego zawodu, jak dietetyk multimedialny. Może dobrze by było, żeby istniał ktoś, kto by chronił „naiwnych” użytkowników Internetu. Do referatu pana Hermy – dla językoznawców bardzo ciekawe jest odkrycie czegoś, co pan określił jako interlingua. Ale z drugiej strony wiemy, że na poziomie leksykalnym współczesny świat, niezależnie od regionu, operuje w bardzo dużym stopniu bardzo wieloma internacjonalizmami, a zatem pokrewieństwa, przynajmniej leksykalnie, nie są zaskakujące. Chciałam zapytać, czy ta interlingualność wykryta przez wrzucenie języka gruzińskiego do języków bliskich przynajmniej geograficznie, używanych w Indiach, dotyczyła też poziomu gramatycznego, czy tylko i wyłącznie poziomu leksykalnego. I jeszcze do wystąpienia doktora Troszyńskiego – szkoda, że nie pokazał pan przykładów tekstów, które brzmią niewinnie, a jednak zostaną rozpoznane przez komputer jako „nienawistne” (zapytam też humorystycznie, czy ten test przeszłyby takie dzieła, jak na przykład *W pustyni i w puszczy* Henryka Sienkiewicza).

**Anna Szymanek:** Jestem tłumaczką, zajmuję się obecnie tłumaczeniami pisemnymi, z ustnymi miałam do czynienia w przeszłości. Być może ciekawe byłoby zbadanie wpływu coraz powszechniejszego stosowania tłumaczeń maszynowych na jakość języka w ogóle. Z doświadczenia własnego wiem, że maszyny, które tłumaczą, korzystają z najczęstszych rozwiązań, najpowszechniejszych i często najbardziej sztamponowych. I powielają również nie najlepsze ludzkie tłumaczenia, przez co mamy takie konstrukcje jak *ramy wieloletnie dotyczące problematyki związanej z kwestią*, które w zasadzie nie mają żadnego sensu, mimo że oryginał jakiś sens był.

**Piotr Zbróg:** Bez trudu można wymienić przykłady negatywnego wpływu tłumaczeń na język polski. Na przykład *hard evidence*. Zamiast mówić po polsku *mocne*,

*niezbite dowody*, przyzwyczailiśmy się do kalki językowej – *twarde dowody*. Nawet tłumacz Google'a przekłada tę frazę m.in. właśnie w ten sposób. Czasem niefortunny przekład nie znajduje odzwierciedlenia w słownikach. W filmie „Dexter” wyrażenie *crime scene* tłumaczono dosłownie jako *scena zbrodni*. Po polsku powie się jednak *miejsce zbrodni*. I chociaż w wielu serialach pojawiała się *scena zbrodni*, Google podaje, że *crime scene* to *miejsce zbrodni*. Zatem jesteśmy w połowie drogi do osiągnięcia sukcesu. Pytanie do doktora Troszyńskiego. Mówił Pan o metodologii szkolenia, tworzenia algorytmów w rozpoznawaniu mowy nienawiści. Nam jako językoznawcom przydałyby się faktyczne przykłady działania tego algorytmu. To jest bardzo istotne, bo wtedy moglibyśmy lepiej zrozumieć, na czym polega trening maszyny elektronicznej. Na przykład czy da się zaprojektować algorytm do oskarżania tego, kto gest pozdrowienia hitlerowskiego zamienia na „pięć piw proszę”? Niemal każdy jest w stanie to rozpoznać, ciekawe czy poradziłaby sobie z tym maszyna. Gdybyśmy otrzymali konkretne przykłady, moglibyśmy dostrzec, czego nie uznano za mowę nienawiści, mimo że sprawiało takie wrażenie, a co uznano, choć na przykład pierwsze wrażenie wcale nie było oczywiste.

## ODPOWIEDZI DYSKUTANTÓW

**Marek Troszyński:** Moje wystąpienie, poza tym, że poświęcone było metodologii badań, to przedmiotowo poświęcone było mowie nienawiści. Zatytułowałem je „Czy komputer rozpozna hejtera?” – użyłem słowa *hejter* zgodnie z uzusem. Jasne jest, czym jest hejter, ale samo użycie anglicyzmu trochę osłabia znaczenie tego zjawiska. Wiemy, że w Sieci ludzie stosują mowę nienawiści i mówią o sobie „jestem hejterem i co z tego”. Gdyby ktoś powiedział „jestem nienawistnikiem”, nadałby temu zjawisku zupełnie inny ciężar gatunkowy. *Hejter* brzmi niemal jak *skejter*, kojarzy się z czymś miłym, pozytywnym – bycie hejterem to coś słabszego niż bycie nienawistnikiem.

Odpowiadając na pytanie o wyniki projektu – celem wystąpienia było pokazanie, w jaki sposób narzędzia automatycznego przetwarzania języka naturalnego mogą współpracować z badaniami społecznymi. Ale rzecz jasna to nie był tylko projekt metodologiczny, choć to była jego istotna część. Od kilku lat prowadzę różne badania, na różne sposoby starając się opisać zjawisko mowy nienawiści w polskim Internecie, w ostatnich dwóch latach skupiam się na zjawisku mowy nienawiści w kontekście migrantów z Ukrainy, Ukraińców mieszkających w Polsce i Ukrainy samej, jako państwa (odsylam państwa do dwóch raportów). Wspólnie ze Związkiem Ukraińców w Polsce zadaliśmy sobie trud przeanalizowania polskich mediów społecznościowych pod kątem tego, w jaki sposób formułowane są opinie dotyczące Ukraińców i Ukrainek w Polsce. Pierwszy raz od II wojny światowej społeczeństwo polskie przestało być homogeniczne – mamy realnie istniejącą mniejszość w Polsce, realnie istniejącą w tym sensie, że przeciętny człowiek spotyka się z przedstawicielami tej mniejszości. W związku z tym dyskurs o Ukraińcach (w przeciwieństwie do dyskursu antysemitckiego czy antyislamskiego, które nie mają odniesienia przedmiotowego) opisuje realne doświadczenia ludzi. Poza tym, że staramy się opisać, jak wiele jest komentarzy, jaka część z nich ma negatywny wydźwięk, a jaka ma wydźwięk pozy-



tywny, to również staramy się podzielić te teksty przedmiotowo, jeśli chodzi o obszar, którego dotyczą. Udało nam się pokazać w badaniach pewną prawidłowość – w tekstach, których tematyka odnosi się w jakikolwiek sposób do historii, dotyczy tematów historycznych, odsetek wypowiedzi niechętnych, nienawistnych, negatywnych jest znacząco wyższy niż w innych tekstach. Dla odmiany w tekstach, które zaklasyfikowaliśmy jako teksty dotyczące ekonomii, w tym również rynku pracy, odsetek wypowiedzi negatywnych jest znacząco niższy. To jest proporcja mniej więcej 30% dla negatywnych tekstów dla tematu ekonomia (w tym rynek pracy) i ponad 50% dla tematu historia. Przyporządkowanie do dziedziny jest wykonane metodami automatycznymi, z wszystkimi zastrzeżeniami, o których mówiłem. Na konferencjach staram się nie prezentować rzeczywistych tekstów źródłowych, żeby nie epatować niepotrzebnie mową nienawiści.

Drugie pytanie dotyczyło przestrzeni publicznej, roli portali internetowych i tego, w jaki sposób można by tę przestrzeń publiczną uporządkować. Padło stwierdzenie, że wystarczy przestrzegać prawa. Jak najbardziej jestem za tym, że trzeba przestrzegać prawa, tylko żeby pokazać, że dany tekst łamie polskie prawo, w szczególności Kodeks karny, trzeba przeprowadzić długi i żmudny proces łącznie z procesem sądowym, i dopiero jeśli sąd wskaże, że rzeczywiście został naruszony na przykład artykuł 256 Kodeksu karnego (mówiący o nawoływaniu do nienawiści), to wtedy mamy do czynienia ze złamaniem prawa. Mamy też ustawę o świadczeniu usług drogą elektroniczną, która to ustawa, na wzór amerykańskiego przepisu z lat 90., zwalnia pośredników internetowych z odpowiedzialności za słowo, w przeciwieństwie oczywiście do wszystkich innych mediów, z wyjątkiem sytuacji, kiedy ten pośrednik zostanie w prawidłowy sposób poinformowany, że naszym zdaniem na jego portalu, na jego komputerze, czy gdziekolwiek indziej znajduje się teść, która naszym zdaniem jest nawoływaniem do nienawiści, a więc łamie polski kodeks karny. I w praktyce wygląda to tak, że pośrednicy, czyli na przykład duże portale horyzontalne, po otrzymaniu takiej wiadomości po prostu chowają, niekoniecznie kasują, ten konkretny post, on przestaje być publiczny, bo z ich perspektywy to nie jest gra warta świeczki, żeby procesować się o ten jeden post. Ale z drugiej strony wiemy, że całkowite usuwanie mowy nienawiści jest wbrew interesowi właścicieli dużych portali i wbrew interesowi pośredników. I dla jasności, mówię o mowie nienawiści w wąskim rozumieniu, czyli tych wszystkich aktach komunikacyjnych, które są nawoływaniem do nienawiści, w szczególności do nienawiści wobec mniejszości. Inaczej mówiąc, to jest kwestia tylko i wyłącznie woli portali, tylko i wyłącznie nakładów portali. Tłumaczenia, które najczęściej brzmią: to jest niemożliwe, tych komentarzy jest zbyt wiele, są tłumaczeniami nieprawdziwymi. Po pierwsze – mamy metody automatyczne, po drugie – można zatrudnić nie dwóch czy czterech moderatorów, tylko na przykład czterdziestu, którzy poradzą sobie ze wszystkimi zgłaszanymi tekstami albo z tymi tekstami, które zostały wstępnie wskazane przez automaty jako budzące wątpliwości, i nie dopuszczą do ich publikacji. Jeśli rzeczywiście serio chcemy dbać o przestrzeń publiczną, to jest rozwiązanie dosyć proste. Choć oczywiście kosztowne i te koszty w moim przekonaniu powinny być przeniesione na biznes, który zarabia na publikacji tych treści, pomimo że uparcie twierdzi, że jest tylko pośrednikiem.

Padło pytanie o patostreaming. To jest rzeczywiście bardzo istotny problem w Polsce, zajmował się nim nawet Rzecznik Praw Obywatelskich. Internet działa

przede wszystkim na młode osoby, a w tym obszarze medialnym normy społeczne działają znacznie słabiej niż w realnym społeczeństwie. Jeśli You Tube czy Google przestanie pozwalać na emisję takich filmów, to łatwiej będzie przekonać młodych ludzi, że pokazywane w nich treści są społecznie niedopuszczalne. Ale jeśli nie, to młodzi ludzie będą je oglądać, bo zło jest atrakcyjne.

I ostatnia rzecz – w sprawie *W pustyni i w puszczy*. Kiedy mówimy o mowie nienawiści w wąskim rozumieniu, to problemem nie są teksty rodzaju *W pustyni i w puszczy*, problemem nie są teksty ironiczne (zresztą z ironią metody automatyczne nie do końca potrafią sobie poradzić). Narzędzia, o których mówiłem, są wykorzystywane do tego, żeby preselekcjonować tekst, żeby wskazać grupę tekstów, które później człowiek przegląda, jeśli chce znaleźć takie, o których warto poinformować pośrednika, że nie powinny się znajdować na portalu, a w przypadku braku reakcji złożyć zawiadomienie do prokuratury, jeśli tekst w jawny sposób nawołuje do nienawiści. Ale raz jeszcze podkreślam, że kiedy mówię o mowie nienawiści, to mam na myśli teksty wprost nawołujące do zachowań, które są bezpośrednio przemocą fizyczną wobec mniejszości, a nie są delikatną aluzją czy szerszym nawiązaniem. Oczywiście, byłoby formalnie ciekawym eksperymentem sprawdzić, w jaki sposób program zaklasyfikowałby tekst *W pustyni i w puszczy*, tylko nie temu te narzędzia służą. Ich zadaniem jest znalezienie w tej wielkiej masie tekstów wypowiedzi, które bezpośrednio nawołują do nienawiści, lub też pokazać różne inne cechy tych tekstów, rozumianych jako korpusy oczywiście, nie jako pojedyncze wypowiedzi.

**Mirosław Filiciak:** Na wstępie ważna uwaga – zmieniły się dzisiaj obiegi kultury: nie jest tak, jak było kiedyś, że osoby o wysokim kapitale kulturowym konsumują niewielką ilość treści ze szczytu piramidy hierarchicznej (na przykład słuchają muzyki klasycznej), tylko piramida się odwróciła, bo mamy dostęp do edukacji, mamy dostęp do treści. W związku z tym osoby z wysokim kapitałem kulturowym obecnie coraz częściej konsumują szeroko, na przykład ironicznie słuchają disco polo. Na marginesie – to je odróżnia od osób, które słuchają tylko disco polo.

Odpowiadając na pytanie dotyczące dietetyka albo dietetyczki medialnej (swoją drogą, to nie jest mój autorski pomysł) – to jest bardzo dobry koncept, tylko utopijny. Mam na myśli utopijną fantazję o oddolnym Internecie, który będzie inkluzywną przestrzenią deliberacji. Prawdopodobnie większość z nas zna kogoś, kto się bardzo dobrze zna na jakimś rodzaju kina czy muzyce, i te osoby będą nam podpowiadać, co obejrzeć lub czego słuchać. Problem polega na tym, że kiedy dużą część tej wymiany komunikacyjnej filtrują media społecznościowe, to te osoby nieczęsto się przebijają przez te filtry, ponieważ algorytmy działają jednak w ramach podstawowego paradygmatu, że jeśli sobie kupiłem jakieś buty, to prawdopodobnie będę chciał kupić znowu podobne buty, trochę inne, i dostaję tak naprawdę więcej tego samego. Więc dostaję to, co już polubiłem, co już skomentowałem, z czym wszedłem w jakąś interakcję. Myślę, że właśnie to jest problem – zamykanie w bańkach powoduje, że podsuwa się nam więcej tego samego, nie rozszerza się nam diety, lecz utwierdza nas się w przekonaniu, że świetnie wybieramy i dostajemy więcej tego samego. Chociaż oczywiście są też modele alternatywne, mamy też firmy takie jak Apple, które każą sobie płacić – najczęściej słono – za swoje usługi i występują w modelu kuratorskim. Ich model biznesowy polega na tym, że u nas nie dostaniesz za darmo, u nas dostaniesz drożej niż gdzie indziej, ale my dla ciebie wybierzemy to, co jest odpowiednie. Logika tego

wyboru wydaje się inna niż ta, na której by nam być może zależało. Oczywiście znowu możemy mówić, że kapitał kulturowy, edukacja są czymś, co mogłoby stanowić jakąś szczepionkę, jednak nie należy być w tej kwestii zbytnim optymistą.

Można sobie zadać pytanie o to, jak się bronić przed *fake newsami*. Myślę, że – nie bagatelizując roli edukacji, która jest warunkiem niezbędnym, ale nie wystarczającym – szybko i skutecznie można to uczynić jedynie za pomocą odgórnych regulacji, ale ich wprowadzenie mogłoby być bardzo trudne, między innymi dlatego, że są to kwestie dla znacznej części społeczeństwa abstrakcyjne, specjalistyczne. Kiedyś rozważano pomysł podatku cyfrowego, wystarczyło jednak kilka zdań wiceprezydenta USA i już podatku cyfrowego nie będzie. Wolność Internetu to mit, bo media o takiej sile oddziaływania zawsze były regulowane odgórnie. Inna sprawa, że mamy już kraje, które dość dobrze sobie radzą z regulowaniem Internetu, to są przede wszystkim Chiny i Rosja – i tutaj pojawia się dość łatwe narzędzie dla firm, które nie chcą być dokładniej kontrolowane. Wystarczy zapytać: czy chcemy iść w tym kierunku?

Wreszcie jest w tym wszystkim kłopot związany z tym, że to są usługi, z których korzystamy i z których lubimy korzystać za darmo. Robiłem kiedyś badanie, w którym pytaliśmy ludzi najpierw o kwestie związane z wartościami, na przykład czy prywatność jest ważna. Dla wszystkich jest ważna i to są absolutnie fundamentalne sprawy. A potem pojawiały się pytania o sprawy niemożliwe do zrealizowania, ale chodziło nam o ćwiczenie myślowe: ile byłabyś/byłbyś w stanie zapłacić za Facebooka, gdyby w zamian za tę opłatę nie gromadzono twoich danych? Respondenci podawali kwoty oscylujące wokół złotówki czterdziestu groszy miesięcznie – jak na coś, co jest dla nas fundamentalnie ważne, cena była niewysoka. Problem polega na tym, że ludzie nie uświadamiają sobie tego, że nie tylko są klientami firm, które dają nam coś za darmo, ale są też darmowymi pracownikami, bo to, co wpisujemy, napędza zyski tych firm. Jesteśmy też zasobem, mówi się często, że nową ropą są dane i my jesteśmy nowym polem naftowym, z którego ktoś zasysa dane, nie pytając o zgodę – a już na pewno nie pytając w sposób, który uświadamia nam wagę tej decyzji. I wreszcie jesteśmy też towarem, ponieważ nasze dane są sprzedawane innym firmom. Jednocześnie wygodnie jest mieć darmowe usługi i dostać rekomendacje pod nos. Wydaje mi się, że problemem jest kwestia języka, znalezienie takiego języka, którym na tematy trudne, specjalistyczne można by było rozmawiać.

**Mariusz Herma:** Padło tu słowo *wygoda* i myślę, że ono jest najważniejsze. W Internecie idziemy na skróty i korzystamy z tego, co jest wystarczająco dobre, a jest pod ręką, jest wygodne – nawet jeśli nie jest najlepsze. W zdecydowanej większości ludzie korzystają więc z tego, co jest darmowe, a zarazem wystarczająco dobre. Trzeba przyznać, że usługi Facebooka czy Twittera są bardzo dobre. Przerastają oczekiwania, często je wyprzedzają.

Google Translate czy Bing Microsoft Translator nie są jedynymi tak dobrymi tłumaczami na świecie. Chińczycy też są znakomici, bo oni bardzo zainwestowali w sztuczną inteligencję, rząd i korporacje działają tu wspólnie. O ile zachodnie koncerny mają pewne ograniczenia, szczególnie Unia Europejska nakłada kolejne, to w Chinach jest pełna wymiana danych. Kamery zainstalowane na ulicach karmią algorytmy niesamowitą liczbą danych. A w tłumaczeniach Chińczycy deklarują, że mają już 95% tego, co chcą osiągnąć – tak jest w ich aplikacji WeChat, która jest

w zasadzie Internetem w Chinach i oferuje dziesiątki różnych usług, jak gdyby łącząc Google'a, Facebooka, Apple'a, Amazona i Allegro.

Przejdziemy teraz do psucia języka przez tłumaczenia. Była dosyć głośna afera w Chinach, bo algorytm nauczył się, że słowo *czarnoskóry*, *Afroamerykanin*, *czarny* może tłumaczyć w najgorsze możliwe sposoby – jako *czarnuch*. Wywołało to protesty, serwis się zorientował i poprawił to ręcznie – to był kolejny krok do jakości tłumaczeń. Na pewno ta jakość jest inaczej rozumiana przez użytkowników Google'a, przez Google'a i przez tłumaczy, którzy to obserwują. Google'owi i większości użytkowników takich serwisów jak Google Translate nie zależy na jakości transliteracji, tylko na komunikatywności. Na „wystarczająco dobrym”. Ale Google wprowadziło możliwość poprawiania tłumaczeń. Jeśli państwo użyją aplikacji do jakiegoś tłumaczenia, to można to wyedytować i on to zapisze, zapamięta, że ktoś zrobił tę poprawkę. Gdy wystarczająca liczba osób coś takiego zasugeruje, maszyna nauczy się to robić lepiej, a przy niektórych popularnych frazach pojawi się znaczek „ptaszka” wskazujący, że to zostało zweryfikowane przez społeczność. Tak więc ciekawe jest, że on się uczy od ludzi, coś proponuje, a potem ludzie to znowu akceptują i poprawiają.