

Paweł Trzaskowski, Obserwatorium Etyki Słowa UW

Nie sposób dialogować z pięścią. O skuteczności zewnętrznej moderacji w walce z nieetycznymi komentarzami internetowymi

The impact of user moderation on unethical speech in the comment sections of websites.

ABSTRAKT:

W ARTYKULE TYM ZAPREZENTOWANE SĄ WYNIKI BADANIA, KTÓRE MIAŁO NA CELU PRZEANALIZOWAĆ WPŁYW ZEWNĘTRZNEJ MODERACJI WYPOWIEDZI INTERNETOWYCH NA KSZTAŁT INTERNETOWEGO POLILOGU, A W SZCZEGÓLNOŚCI NA ZACHOWANIA TYCH UŻYTKOWNIKÓW, KTÓRZY ZAMIESZCZALI KOMENTARZE NIEETYCZNE. BADANIE ZOSTAŁO PRZEPROWADZONE NA TRZECH POPULARNYCH POLSKICH PORTALACH. BADACZE STOSOWALI ROZMAITE STRATEGIE DYSKURSYWNE, ABY NAKŁONIĆ KOMENTUJĄCYCH DO ZANIECHANIA STOSOWANIA NIEETYCZNEGO JĘZYKA. WSZYSTKIE PRÓBY ZAKOŃCZYŁY SIĘ NIEPOWODZENIEM. DOŚWIADCZENIA Z TEGO EKSPERYMENTU PODPOWIADAJĄ PRZYJĘCIE ODMIENNEJ STRATEGII W WALCE Z INTERNETOWYM HEJTEM. AUTOR ARTYKUŁU SUGERUJE, ŻE NIEETYCZNE WYPOWIEDZI NALEŻY TRAKTOWAĆ NIE JAKO AKT KOMUNIKACJI, LECZ JAKO AKT WANDALIZMU. PRZYPOMINA POPULARNĄ W KRYMINOLOGII TEORIĘ ROZBITYCH OKIEN I POSTULUJE WYKORZYSTANIE JEJ ZAŁOŻEŃ DO WALKI Z NIEETYCZNĄ MOWĄ W SIECI.

SŁOWA KLUCZOWE:

HEJT, HATE, MOWA NIENAWIŚCI, ETYKA SŁOWA

ABSTRACT:

THE AIM OF THIS PAPER IS TO PRESENT THE RESULTS OF A RESEARCH ON UNETHICAL SPEECH CONDUCTED IN 2014. THE MAIN OBJECTIVE OF THIS RESEARCH WAS TO MEASURE THE IMPACT THAT USER MODERATION HAS ON THE STYLE AND LANGUAGE OF OTHER USERS' COMMENTS IN THE COMMENTS SECTION OF THREE POPULAR POLISH WEBSITES. SEVERAL DISCOURSE STRATEGIES WERE EMPLOYED TO TRY TO CONVINCING THE USERS WHO USED UNETHICAL SPEECH TO ABANDON SUCH A LANGUAGE; HOWEVER, NONE OF THEM WORKED. IN THE LIGHT OF THIS UNSUCCESSFUL ATTEMPT THE RESEARCHER SUGGESTS ADOPTING A NEW APPROACH TO INTERNET HATE – INSTEAD OF TREATING IT AS AN ELEMENT OF COMMUNICATION IT SHOULD BE REGARDED AS VANDALISM AND REMEDIED ACCORDINGLY. THE AUTHOR BORROWS THE BROKEN WINDOWS THEORY FROM THE FIELD OF CRIMINOLOGY AND SUGGESTS EMPLOYING IT IN ORDER TO FIGHT INTERNET HATE BY IMMEDIATELY ERASING SUCH COMMENTS.

KEYWORDS:

ETHICS OF LANGUAGE, HATE, HATE SPEECH

Mowa nienawiści, hejt, *trolling*, *flejmy*, *stalking*, *griefing* i cyberprzemoc to niektóre ze zjawisk istniejących dzięki jeszcze nieuregulowanej, swobodnej komunikacji internetowej. Część z nich to nowe zachowania językowe, niektóre zostały niedawno wyodrębnione i nazwane. Brakuje na razie dokładnych opracowań naukowych, które umożliwiłyby wskazanie cech dystynktywnych poszczególnych typów wypowiedzi internetowych, które naruszają zasady etyki słowa. Nie ma też na razie jednego, ogólnego terminu dla wirtualnej agresji językowej¹. Dlatego też prace mające na celu opisanie i rozróżnienie poszczególnych typów nieetycznych wypowiedzi w internecie są konieczne, zwłaszcza w obliczu postępującego rozluźnienia norm etycznych z jednej strony i nieuchronnego „cywilizowania” i „okiełznywania”, sieci z drugiej. W 1996 roku John Perry Barlow stworzył Deklarację Niepodległości Cyberprzestrzeni, w której postulował utopijną wizję sieci rządzonej przez użytkowników, niezależnej od przemysłowych gigantów². Dziś, ponad dwie dekady później, za organizację internetu zaczynają brać odpowiedzialność te firmy, które przez ten czas budowały swoją potęgę.

Jednym z takich procesów „porządkujących” jest przedsięwzięcie Microsoftu, Facebooka, Twittera i Google’a ogłoszone w maju 2016 roku. Cztery największe firmy przemysłu wirtualnego w porozumieniu z Komisją Europejską zobowiązały się do usuwania informacji, które szerzą nienawiść, w ciągu 24 godzin po ich zamieszczeniu³. Oznacza to konieczność wyznaczenia granicy między mową nienawiści, która podlega penalizacji, a pozostałymi negatywnymi komentarzami. Należy jednak pamiętać, że treści zamieszczane na stronach Facebooka czy Twittera prawnie należą do właścicieli stron, którzy sami mogą określić, które z wypowiedzi spełniają ich definicję mowy nienawiści. Dlatego też jednym z ważnych zadań współczesnego językoznawstwa jest wypracowanie narzędzi, które umożliwiłyby obiektywne, wolne od różnych korporacyjnych wizji poprawności politycznej oddzielenie wypowiedzi podszytych nienawiścią od tych, które motywowane są strachem, wstydem, zazdrością, oportunistycznym czy po prostu zwyczajną złośliwością. Takiego rozróżnienia trzeba dokonywać z rozwagą, ponieważ celowo niesłuszne oskarżenie kogoś o stosowanie mowy nienawiści również jest dyskursywną strategią manipulacyjną i narusza zasady etyki słowa⁴.

¹ Mogłoby się zdawać, że taką funkcję pełni czasami określenie „hejt”, chociaż etymologia tego słowa sugeruje, że pochodzi ono od potocznego wyrażenia „to hate on”, które oznacza niechęć do kogoś motywowaną zawiścią bądź zazdrością (Trzaskowski 2016). Tym samym termin „hejt”, podobnie jak np. „mowa nienawiści”, służy raczej do opisu aktów agresji językowej powodowanych przez jedno, specyficzne uczucie, niewykluczone jednak, że stanie się słowem służącym do opisu ogółu nieetycznych komentarzy internetowych.

² J.P. Barlow, *A declaration of the Independence of Cyberspace*, „Electronic Frontier Foundation” z 08.02.1996 r., <https://www.eff.org/cyberspace-independence> (dostęp 22.10.2017 r.).

³ European Commission, *European Commission and IT Companies announce Code of Conduct on illegal online hate speech*, „European Commission - Press release” 31.05.2017 r.

⁴ P. Trzaskowski, *Różne sposoby mówienia o mowie nienawiści w polskich dziennikach*, „Studia Pragmatyngwistyczne” 2015 VII, s. 69-83.



Równie ważne jest stworzenie metod, które pozwalałyby na skuteczne przeciwdziałanie wszystkim typom nieetycznych wypowiedzi. Komentarze wulgarne, obelżywe, poniżające i wykluczające nie muszą być zamieszczane z powodu nienawiści, nie oznacza to jednak, że nie są szkodliwe.

Równie ważne jest stworzenie metod, które pozwalałyby na skuteczne przeciwdziałanie wszystkim typom nieetycznych wypowiedzi. Komentarze wulgarne, obelżywe, poniżające i wykluczające nie muszą być zamieszczane z powodu nienawiści, nie oznacza to jednak, że nie są szkodliwe. Wirtualna rzeczywistość, czyli zgodnie z definicją słownikową ta, która jest „wykreowana na ekranie komputera, ale tak realistyczna, że wydaje się rzeczywista”⁵ przeobraziła się już w nową platformę uczestniczenia w życiu. Wraz z rozwojem mediów społecznościowych i wzrastającym naciskiem na samoidentyfikację oraz nawiązywanie więzi społecznych w sieci, internet zatracą swoją pierwotną funkcję ogólnodostępnego repozytorium informacji⁶. Zamiast tego stał się forum prywatnych opinii, często wyrażanych językiem wulgarnym, krzywdzącym, zdegradowanym⁷. Te opinie, wbrew przekonaniu o wszechogarniającej sieci, trafiają do stosunkowo wąskiego grona odbiorców⁸ i często mają bardzo mocny wpływ na to, co jeszcze nazywamy „realnym” życiem. Nielubiany uczeń w szkole wraca do domu i jest dodatkowo dręczony na portalu społecznościowym, zazdrosny chłopak wrzuca nagie zdjęcia swojej dziewczyny do internetu, pracownik traci pracę, ponieważ polubił niewłaściwy profil, czyjaś kariera polityczna kończy się po ujawnieniu kompromitujących e-maili. Wirtualny świat nie jest już substytutem realnego świata, lecz splótł się z nim nierozzerwalnie,

⁵ www.sjp.pwn.pl, hasło *Wirtualna rzeczywistość*.

⁶ M. Koza, *Czy znasz ten mem? Pragmatyka i polityka internetowych wspólnot interpretacyjnych*, „Teksty Drugie. Cyfrowa Piśmienność” 2015 z. 3, s. 236-245.

⁷ K. Krzysztofek, *Mowa nienawiści jako przemoc*, „Societas/Communitas” 2010 z. 2, s. 22-26.

⁸ To zjawisko określa się mianem „echo chamber” (ang. „kółko wzajemnej adoracji”) i polega na dostarczaniu użytkownikowi tylko tych treści, które odpowiadają jego gustom i zainteresowaniom w oparciu o algorytmy takie jak newsfeed oraz preferencje znajomych użytkownika. Doskonałym przykładem działania zasady „echo chamber” był eksperyment przeprowadzony przez „The Washington Post” podczas wyborów prezydenckich w Stanach Zjednoczonych w 2016 roku. W czasie rzeczywistym zestawiono ze sobą informacje, które na portalach społecznościowych otrzymywali zwolennicy republikanów i demokratów. Okazało się, że obie grupy otrzymują dwie diametralnie różne wersje rzeczywistości. Wciąż jednak największy wpływ na przekazywanie informacji w „kółku wzajemnej adoracji” mają opinie znajomych, z którymi użytkownik utrzymuje kontakt w internecie (Bakshy, Messing, Adamic 2015: 1130-1132).

stał się jego przedłużeniem. Dlatego też należy troszczyć się o jego jakość, w tym o język, którego się w nim używa.

Celem tego artykułu jest zaprezentowanie wyników analizy przeprowadzonej na forach internetowych w ramach prac prowadzonych przez Obserwatorium Etyki Słowa⁹, zespołu badawczego Uniwersytetu Warszawskiego kierowanego przez prof. Annę Cegiełę. Przeprowadzono badanie, które polegało na określeniu wpływu zewnętrznej moderacji, czyli tej inicjowanej przez samych użytkowników danego portalu, na akty słowne, które naruszają zasady etycznej komunikacji. Badano wpływ zewnętrznej moderacji bezpośredniej, czyli takiej, która wymaga wejścia w interakcję z komentującymi (w przeciwieństwie do moderacji zewnętrznej pośredniej, która polega na zgłaszaniu kontrowersyjnych postów do moderatorów stron). Badanie przeprowadzono w październiku 2014 roku.

Przebieg badania

Do analizy wybrano trzy opiniotwórcze portale o odmiennej orientacji światopoglądowej: wyborcza.pl, fronda.pl i onet.pl, kierowano się przy tym poczytnością oraz rozpoznawalnością stron¹⁰. Z każdego portalu do dalszego etapu badania wybrano po trzy artykuły, ważna była przy tym data opublikowania danego tekstu oraz liczba komentarzy, które zgromadził – kolejny etap badania wymagał, żeby analizowane artykuły były możliwie jak najnowsze oraz stosunkowo popularne. Następnie sprawdzono, jaką część wszystkich komentarzy stanowią wypowiedzi, które naruszają zasady etyki słowa. Za kryterium przyjęto model etycznej komunikacji zaproponowany przez Cegiełę¹¹. Osobie, o której się mówi, Drugiemu, należą się takie same prawa jak temu, który mówi. Brak poszanowania ludzkiej autonomiczności, pozbawienie człowieka jego wielowymiarowości oraz ludzkiego wymiaru są wg Cegieły podstawowymi wykroczeniami wobec etyki słowa, realizowanymi w takich strategiach dyskredytujących jak etykietowanie, stygmatyzacja, stereotypizacja, dehumanizacja i depersonifikacja. Tym samym każdy komentarz, w którym zastosowano którąkolwiek z powyższych strategii, był klasyfikowany jako nieetyczny.

Analiza wykazała, że większość komentarzy zamieszczonych pod artykułami miała wydźwięk negatywny, wiele z nich miało charakter nieetyczny. Poza nieetycznością samych wypowiedzi częstym zjawiskiem była ich niekompletność, błędy logiczne, ortograficzne, problemy z zapisem (rozstrzelenia, cały tekst zapisany kapitalikami). Komentujący rzadko wchodzili w złożone konwersacje, większość wiadomości miała

⁹ www.oes.uw.edu.pl.

¹⁰ W momencie przeprowadzania badania w ankiecie Millward Brown onet.pl został wskazany jako najbardziej rozpoznawalny portal w Polsce. Wyborcza.pl została wybrana do badania ze względu na zajęcie pierwszego miejsca w poczytności polskich dzienników (bez uwzględnienia tabloidów). Natomiast fronda.pl zwróciła uwagę badaczy nie tylko poczytnością (choć udział komentujących jest w porównaniu do Onetu i Wyborczej stosunkowo skromny), lecz również ze względu na język zamieszczanych tam artykułów, który jest wyjątkowo upraszczający i potępiający.

¹¹ A. Cegieła, *Słowa i ludzie, wprowadzenie do etyki słowa*. Warszawa 2014.

albo charakter jednostkowy, albo była słowną utarczką między dwoma użytkownikami. Popularnym zjawiskiem było również przybieranie przez komentujących prowokujących, agresywnych bądź obraźliwych pseudonimów. Zaobserwowano również różne metody wewnętrznej moderacji¹². Po rozpoznaniu przystąpiono do kolejnego etapu badania, czyli próby zewnętrznej moderacji bezpośredniej oraz jej wpływu na kształt internetowego polilogu.

Badacze przyjęli neutralne pseudonimy oraz podjęli próbę nawiązania kontaktu z tymi użytkownikami, którzy zamieścili nieetyczne komentarze. Starali się nakłonić ich do prowadzenia etycznego dialogu. Stosowali środki perswazji słownej, takie jak upomnienie, prośba, nagana, wytknięcie błędu logicznego. Próbowali nawiązać merytoryczną rozmowę z innymi użytkownikami w celu zagłuszenia nieetycznych wypowiedzi innych. Prowadzili konwersację między sobą, aby stymulować etyczne i konstruktywne wypowiedzi pozostałych użytkowników. Zwracali się również z prośbą o pomoc do innych komentujących. Wszystkie próby podejmowane przez badaczy pozostały jednak niezauważone przez pozostałych użytkowników portali, nie wywoływały żadnej reakcji. Nie udało się tym samym wskazać najbardziej skutecznej formy perswazji w moderacji zewnętrznej, ponieważ żadna z przedstawionych strategii nie wpłynęła na zachowania pozostałych uczestników internetowego polilogu.

Wnioski z badania

Niepowodzenie zewnętrznej moderacji prowadzonej w ramach badania pozwoliło na sformułowanie pewnych wniosków dotyczących charakteru nieetycznych komentarzy internetowych.

Po pierwsze, brak efektów moderacji był z pewnością po części rezultatem niewystarczającej liczby głosów, które zostały wprowadzone do dyskusji przez badaczy. Oznacza to, że przedstawiciele grupy starającej się o zaprowadzenie porządku musi być stanowczo więcej niż tych, którzy wprowadzają chaos i agresję. Osiągnięcie takiego stanu wydaje się na razie niemożliwe – większość wartościowych głosów nie pojawia się w komentarzach internetowych, ponieważ nie jest to forma, która pozwala na konstruktywną wymianę opinii. Dział komentarze służy tylko wyrażaniu opinii i emocji, a nie dyskusji. Ci z użytkowników, którzy chcieliby dodać wartościowe przemyślenia w komentarzach często z tego rezygnują właśnie ze względu na ton pozostałych wypowiedzi, atmosferę na forum i brak możliwości zaistnienia rzeczowej dyskusji.

Po drugie, udało się wskazać na pewne cechy dystynktywne komentarza internetowego. Afektywność, często nawet wulgarność tych aktów słownych, połączona ze wspomnianą już wcześniej niestarannością zapisu oraz licznymi błędami w samej strukturze wiadomości świadczy o efemeryczności tej formy wypowiedzi. Celowe jest tutaj użycie

¹² Niektóre metody moderacji sprawiały wrażenie, że nie mają na celu zapanowanie nad dyskusją, lecz wprowadzenie jeszcze większego chaosu. Na przykład komentarze na portalu onet.pl, zdecydowanie najbardziej popularnego w kontekście aktywności użytkowników, były bardzo często usuwane, bądź przenoszone, zarówno na górę jak i na sam dół listy komentarzy, właściwie bez uzasadnienia. Dopuszczano również wulgaryzmy, co raczej nie zdarzało się na pozostałych portalach, które były badane.



Nieprzemyślany, agresywny czy nawet wulgarny komentarz jest wyraźnie akceptowalny społecznie (w myśl zasady, że „tak się przecież w internecie mówi”). Internet pozwala na tę agresję, na utrwalenie frustracji przez nadanie zjawiskom prywatnym rangi publicznej oraz przez nadanie zjawiskom ulotnym cech trwałych.

słowa „wypowiedzi”, nie „komunikacji”, ponieważ zachodzi przypuszczenie, że komentarze internetowe, zwłaszcza te nieetyczne, nie mają wiele wspólnego z komunikacją rozumianą jako akt przekazania informacji między nadawcą i odbiorcą. Wydaje się raczej, że celem tych tekstów nie jest też perswazja, nawet w wulgarniej formie, lecz uzewnętrznienie negatywnych emocji komentującego. Świadczy o tym również jednorazowość takich wpisów, utrata zainteresowania komentującego swoim komentarzem w chwili zamieszczenia go, oraz brak reakcji na próby nawiązania kontaktu przez innych użytkowników. Zamieszczanie nieetycznych wpisów może też być przejawem celowej, złośliwej działalności, w której irytacja i złość pozostałych użytkowników jest gratyfikacją dla tych, którzy takie komentarze publikują. Przybieranie przez komentujących prowokujących komentarzy może również być rozpatrywane jako element subwersywnej gry, którą prowadzą¹³. Jednak bez względu na motywację w obu przypadkach takie wypowiedzi stają się werbalnymi substytutami agresji fizycznej, które przychodzą łatwiej, ponieważ są anonimowe i wciąż bezkarne. Sądzę również, że sytuacja anonimowości i działania w grupie zwalnia z poczucia winy, ponieważ komentujący zazwyczaj nie widzi skutków swoich słów, więc nie musi myśleć o tym, czy zrobił coś złego¹⁴. Nieprzemyślany, agresywny czy nawet wulgarny komentarz jest wyraźnie akceptowalny społecznie (w myśl zasady, że „tak się przecież w internecie mówi”). Internet pozwala na tę agresję, na utrwalenie frustracji przez nadanie zjawiskom prywatnym rangi publicznej oraz przez nadanie zjawiskom ulotnym cech trwałych. Gatunek wypowiedzi, jakim jest komentarz internetowy, ułatwia, zachęca wręcz do tworzenia wypowiedzi skrajnie emocjonalnych, silnie ekspresywnych, nacechowanych negatywnie. Natomiast brak konsekwencji, zarówno prawnych, jak i osobistych, tworzenia nieetycznych wypowiedzi wpływa na powszechność tego zjawiska.

¹³ Subwersywna gra może być trudna do odróżnienia od wyrachowanej gry - zazwyczaj politycznej. Nie jest tajemnicą, że zorganizowany hejt internetowy stał się już jedną ze strategii wykorzystywanej do wpływania na wyborców.

¹⁴ Aczkolwiek można zauważyć przyrost negatywnych (a w tym nieetycznych) komentarzy na portalach, na których użytkownicy publikują komentarze pod swoim imieniem i nazwiskiem. To rozprężenie można tłumaczyć zjawiskiem zaraźliwego sąsiedztwa, które opisuje w dalszej części tekstu.

Teoria rozbitych okien

Wymienione tu cechy pozwalają na porównanie nieetycznych komentarzy internetowych do aktów wandalizmu. Hejterzy, podobnie jak wandy, którzy malują graffiti¹⁵ na murach, zostawiają po sobie ślad, który ma negatywny wpływ na innych użytkowników przestrzeni publicznej. Środki językowe okazują się niewystarczające do przeciwdziałania, można więc spróbować rozpatrywać komentarz nie jako akt komunikacji, lecz jako akt wandalizmu i zastosować wobec niego strategię wypracowaną na polu kryminologii. W 1982 sformułowano teorię rozbitych okien¹⁶ (Kelling, Wilson 1982). Autorzy tej teorii po obserwacji postępującej degradacji przedmieść wielkich amerykańskich miast doszli do wniosku, że zdewastowana przestrzeń publiczna jest kryminogenna sama w sobie. Takie otoczenie komunikuje ludziom, którzy w nim przebywają, że po pierwsze, miejsce nie jest w dostatecznym stopniu monitorowane i kontrolowane przez służby porządkowe oraz samych mieszkańców, po drugie, społeczeństwo, które je zamieszkuje, cechuje anonimowość oraz brak wyraźnie wyznaczonych norm społecznych, po trzecie społeczeństwu temu brak organizacji a liczba obywateli zdolnych do łamania prawa znacząco przewyższa liczbę tych, którzy próbują je stanowić i po czwarte, ludzie zamieszkujący takie miejsce doświadczają uczucia niepokoju, często nawet strachu. W takim środowisku przestępczość jest większa, dochodzi do eskalacji na prawach zaraźliwego sąsiedztwa¹⁷.

Rozwiązania proponowane przez Kellinga i Wilsona polegały na jak najszybszym naprawianiu szkód, skupianiu się na małych problemach zanim eskalują. Przywrócenie porządku w przestrzeni publicznej miało na celu zwiększenie potrzeby praworządności mieszkańców danego obszaru. Teoria Kellinga i Wilsona była szeroko dyskutowana i stała się podstawą dla licznych eksperymentów, które, w większości, potwierdzały ich tezę¹⁸.

Zakończenie

Teorię rozbitych okien można zastosować w walce z nieetycznymi komentarzami. Internetowym odpowiednikiem zamalowywania graffiti bądź wstawiania szyb może być zgłaszanie takich wypowiedzi do usunięcia moderatorom stron. Moderatorzy działają

¹⁵ Należy przy tym wyraźnie odróżnić prymitywne tagi oraz nienawistne napisy od street artu.

¹⁶ G. L. Kelling, J. Q. Wilson, *Broken windows: the police and neighborhood safety*, „Atlantic Monthly” 1982 z. 249 (3), s. 29-38.

¹⁷ Philip Zimbardo demonstrował skuteczność zaraźliwego sąsiedztwa w swoim badaniu z 1969 roku. Dwa samochody, które wyglądały na opuszczone, pozostawiono w dwóch różnych dzielnicach: na trapiącym przestępczością nowojorskim Bronksie oraz w Palo Alto, zamożnej dzielnicy w Kalifornii. Samochód zaparkowany na Bronksie zaczęto szabrować już po kilku minutach od rozpoczęcia eksperymentu, po trzech dniach, kiedy samochód został pozbawiony wartościowych części, rozpoczęła się jego dewastacja. Natomiast samochód pozostawiony w Palo Alto stał nietknięty przez kilka dni do momentu, kiedy Zimbardo sam nie zaczął niszczyć go młotem. Mieszkańcy okolicy przyłączyli się do dewastacji, samochód został zniszczony (Paynich, Hill 2011, s. 59-62).

¹⁸ W oparciu o tę teorię dokonano rewitalizacji m. in. Nikiszowca i Giszowca, dzielnic Katowic, które borykały się z licznymi problemami społecznymi.

jednak uznaniowo, często w interesie portalu zostawiają agresywne komentarze, które generują ruch na stronie. W opozycji do tych strategii niektóre portale uznały ułomność gatunkową komentarzy internetowych i rezygnują z wyświetlania ich w ogóle¹⁹ (Mead 2015), bądź stosują bardzo rygorystyczną selekcję publikowanych komentarzy. Innym rozwiązaniem wydaje się tworzenie aplikacji, które usuwałyby nieetyczne komentarze z poziomu użytkownika w podobny sposób, w jaki usuwają np. reklamy. W dobie personalizowanego doboru otrzymywanych informacji użytkownik mógłby zupełnie zrezygnować z wyświetlania treści komentarzy, które łamią zasady etyki słowa. Takie rozwiązanie mogłoby również znaleźć zastosowanie jako narzędzie kontroli rodzicielskiej – chroniłoby młodych użytkowników przed wulgarnymi treściami, które pojawiają się często obok neutralnych tematów²⁰. Natomiast w szerszej perspektywie należałoby zbadać sam komentarz internetowy jako gatunek wypowiedzi. Czy jego negatywna realizacja jest wypadkową internetowego permissyvizmu i ciemnych zakamków ludzkiej psychiki, czy może też wyznaczniki gatunkowe nakierowują autora komentarza na tworzenie wypowiedzi chaotycznych, emocjonalnych i często właśnie nieetycznych? Czy reguły, które rządzą tym gatunkiem z góry uniemożliwiają konstruktywną rozmowę? Czy komentarz internetowy niszczy przestrzeń komunikacji, którą stanowi internet i przekształca ją w pole agresji? Warto byłoby zastanowić się nad tym, może się bowiem okazać, że wolność słowa wcale nie ucierpi, jeśli w sieci w ogóle zabraknie komentarzy. ■

BIBLIOGRAFIA:

- Bakshy E., Messing S., Adamic L.A., *Exposure to ideologically diverse news and opinion on Facebook*, "Science" 2015, Vol. 348.
- Barlow, J.P., *A declaration of the Independence of Cyberspace*. www.eff.org/cyberspace-independence.
- Cegiela, A., *Słowa i ludzie, wprowadzenie do etyki słowa*, Warszawa 2014.
- European Commission *European Commission and IT Companies announce Code of Conduct on illegal online hate speech*, w: *European Commission - Press release 31 May 2016*, Bruksela 2016.
- Greenberg, A., *Inside Google's Internet justice league and its AI-powered war on trolls*, <https://www.wired.com/2016/09/inside-googles-internet-justice-league-ai-powered-war-trolls/>.
- Kelling G.L., Wilson J. Q., *Broken windows: the police and neighborhood safety*, w: *Atlantic Monthly*, 1982, 249(3), s. 29-38.

¹⁹ D. Mead, *We're replacing comments with something better*, „Motherboard” z 05.10.2015 r., <http://motherboard.vice.com/read/im-on-twitter-too> (dostęp 22.10.2017 r.).

²⁰ Google rozwija już technologię Conversation AI, która wychwytuje nieetyczne wypowiedzi w oparciu o słowa kluczowe (Greenberg 2016). Boryka się jednak z manipulacjami, które mają na celu oszukanie algorytmu tak, aby zaczął traktować słowa neutralne jako nienawistne (Tait 2016).

- Koza M., *Czy znasz ten mem? Pragmatyka i polityka internetowych wspólnot interpretacyjnych*, „Teksty Drugie. Cyfrowa Piśmienność”, 2015, nr 3.
- Krzysztofek K., *Mowa nienawiści jako przemoc*, „Societas/Communitas”, 2010 vol. 2.
- Mead D., *We're replacing comments with something better*, www.motherboard.vice.com/read/im-on-twitter-too.
- Paynich R., Hill B., *Fundamentals of Crime Mapping*, Curry College 2011.
- Tait A. *“Stinking Googles should be killed”: why 4chan is using a search engine as a racist slur*, <http://www.newstatesman.com/science-tech/internet/2016/09/stinking-googles-should-be-killed-why-4chan-using-search-engine-racist>.
- Trzaskowski P., *Różne sposoby mówienia o mowie nienawiści w polskich dziennikach*, „Studia Pragmalingwistyczne” 2015, s. 69-83.
- Trzaskowski P., *Hejtuję nie zawsze znaczy nienawidzę - zmiany w polu językowym nienawiści*. (w przygotowaniu).

O AUTORZE:

mgr Paweł Trzaskowski - pracownik Obserwatorium Etyki Słowa, sekretarz i członek Zespołu Etyki Słowa Rady Języka Polskiego PAN. Doktorant Wydziału Polonistyki Uniwersytetu Warszawskiego, językoznawca i anglista. W swojej pracy naukowej zajmuje się przede wszystkim problematyką nienawiści. Kontakt: etykaslowa@gmail.com