

Język w służbie marketingu. Wpływ technologii na analizę sentymentu

Wprowadzenie

Pojawienie się Internetu wpłynęło na niemal każdy aspekt życia człowieka, w tym przede wszystkim na sposób, w jaki się komunikujemy (Gherghita-Mihaila 2016). Zmianie uległy język, metody i narzędzia komunikacji zarówno prywatnej, jak i oficjalnej, a zwłaszcza biznesowej. W efekcie istotnym zmianom uległ także tzw. *marketing mix* (4P), czyli zespół elementów, za pomocą których można oddziaływać na rynek (Kotler 2000: 9), analizując zachowania i potrzeby klientów. Pojawiło się w związku z tym nowe pojęcie *analiza sentymentu*, które oznacza kierunek badań analizujących ludzkie opinie, oceny, postawy, poglądy oraz emocje wywoływane przez produkty, usługi, organizacje, osoby fizyczne, sprawy, wydarzenia, tematy i ich atrybuty (Zolty 2015). Obecnie do analizy sentymentu wyrażanego przez użytkowników sieci są wykorzystywane technologie, które pozwalają na szybką identyfikację ich opinii i subiektywnych odczuć, ciągle jeszcze mające sporo ograniczeń wynikających często ze specyfiki języka. W swoim artykule postaram się naświetlić, jak z jednej strony ważne są te technologie we współczesnym marketingu, z drugiej zaś pokazać, że coraz ważniejsza staje się współpraca między językoznawcami a twórcami technologii służącej analizie sentymentu wyrażanego w różnorodnych tekstach zamieszczanych w sieci.

Rola marketingu we współczesnym świecie

Jak stwierdziła Bożena Leśniewska, Executive VP B2B Orange Polska, „dziś nie ma już kanałów sprzedaży, są tylko kanały kontaktu”¹. Słowa te dobrze

¹ Cytat z konferencji PayU Lab, kwiecień 2015 (<https://www.youtube.com/watch?v=jtpKO1yfiYo>).

odzwierciedlają zmiany, jakie nastąpiły wraz z rozwojem i popularyzacją Internetu oraz nowych technologii. Istotnej ewolucji podlega zwłaszcza strategia sprzedażowa wielu firm podążających w ślad za klientem. Klient zaś, aktywnie korzystający z narzędzi pozwalających mu na niemal nieograniczony dostęp do informacji niezależnie od pory dnia czy miejsca, w którym się znajduje, zaczął w nowy sposób wyrażać swoje oczekiwania. Przede wszystkim stał się świadomy tego, że ma prawie nieograniczony wybór, wynikający z jednej strony z globalizacji, rozumianej jako ujednoczenie rynków, obracanych na nich towarów (Giddens 2004: 79-80) i rosnącej konkurencji, z drugiej zaś z łatwości porównywania wielu ofert i wybierania tej najlepszej z jego punktu widzenia. Powszechny dostęp do porównywarek cen (rodzaj serwisu WWW, który pozwala na porównanie cen tych samych produktów w różnych sklepach internetowych²) i serwisów pełnych opinii innych użytkowników sprawił zaś, że klient ma świadomość, iż jego pozycja w relacji ze sprzedawcą jest dominująca. To poczucie przekłada się z kolei na jego oczekiwania względem produktu, poziomu obsługi czy samego sprzedawcy. Do oczekiwań, które wzmocniły się najbardziej wraz ze zmianami technologicznymi, należy m.in. szybkość reakcji na potrzeby klienta, w tym niemal błyskawiczna realizacja samego zamówienia. Wszystkie te postawy w sposób istotny przyczyniły się m.in. do powstania nowych modeli sprzedaży.

I tak obok klasycznego modelu jednokanałowego, w którym klienci wchodzi w kontakt z marką poprzez jeden kanał, zaczęły pojawiać się modele takie, jak: *multichannel* (klienci decydują o wyborze preferowanego kanału: *offline*, *online*, *mobile*; poszczególne kanały nie są jednak zintegrowane i działają niezależnie od siebie), *cross-channel* (klient może korzystać z wielu kanałów, w których jest już postrzegany jednostkowo, acz jego doświadczenia zakupowe w kanałach się różnią), *omnichannel* (wszystkie kanały są ze sobą zintegrowane, a doświadczenie zakupowe ujednoczone; w efekcie zacierają się granice między doświadczeniami *online* i *offline*) (Jorij 2012). Wraz z przechodzeniem klienta do kolejnych kanałów i adekwatnie zmieniającą się strategią sprzedaży narzędzia do pozyskiwania klientów i analizy ich zachowań ulegały ewolucji.

Wszystkie opisane wcześniej zmiany postaw i oczekiwań klientów, wraz z upowszechnianiem się modeli sprzedażowych takich jak *omnichannel*, musiały pociągnąć za sobą zmiany także w komunikacji z klientem, a więc w marketingu. Według Kotlera marketing jest procesem planowania i realizacji koncepcji, ustalania ceny, promocji i dystrybucji idei, dóbr i usług w celu dokonania wymiany mającej służyć osiągnięciu celów organizacji i jednostek. Do realizacji tych celów marketing używa instrumentów, które tworzą tzw. *marketing mix*, czyli zespół elementów, za pomocą których możemy oddziaływać na rynek. Najbardziej

² Por. https://pl.wikipedia.org/wiki/Por%C3%B3wnywarka_cen.

popularna w tym kontekście jest koncepcja 4 P – *produkt* (ang. *product*), *cena* (ang. *price*), *dystrybucja* (ang. *place*) i *promocja* (ang. *promotion*) (Kotler, Keller 2012).

Promocja, z całej czwórki marketingowych „P”, została prawdopodobnie najbardziej dotknięta przez zmiany w zachowaniach klientów, spowodowane pojawieniem się Internetu i nowych technologii. Współczesne narzędzia, czyli tzw. *marketing intelligence* (pozwalający na bieżące zbieranie i analizę istotnych dla firm danych rynkowych, a także na badanie najdrobniejszych gestów klienta), w istotny sposób poprawiły możliwości targetowania³ komunikacji z niespotykaną do tej pory precyzją i dokładnością. Spowodowało to nie tylko ciągłe opracowywanie nowych metod docierania do klientów, nowych formatów reklamowych, ale także zwiększenie się częstotliwości, z jaką się one pojawiają.

Dystrybucja (ang. *place*) fizycznego produktu po pojawieniu się Internetu nie zmieniła się istotnie, ale to, jak bardzo marka jest widoczna, oraz to, czy istnieje możliwość jej dotarcia do oczekiwanego klienta w świecie wirtualnym, już tak. Dziś to dotarcie w dużym stopniu zależy od spozycjonowania marki w wyszukiwarce (działanie mające na celu podniesienie poziomu widoczności strony internetowej w wynikach wyszukiwania np. w Google), tak widoczności organicznej, jak i płatnej, w tym także w mediach społecznościowych. Te ostatnie nie tylko same mają istotny wpływ na klientów (Blossom 2009), ale podlegają także szybkiej ewolucji wraz z ciągle pojawiającymi się nowymi rozwiązaniami.

Produkt (ang. *product*) pozostaje kluczowym elementem całej układanki, ale wraz z pojawieniem się nowych kanałów dystrybucji istotnie wzrosło znaczenia tzw. *user experience* (UX, ang. ‘doświadczenie użytkownika’), czyli całości wrażeń, jakich doświadcza użytkownik podczas korzystania z produktu interaktywnego⁴. Sprzedawanie produktu *online* ma swoje mankamenty w związku z brakiem niektórych elementów dobrego doświadczenia zakupowego, takich jak możliwość dotknięcia, sprawdzenia, dokładnego obejrzenia produktu czy uzyskania szybkiej porady lub odpowiedzi od sprzedającego. Te ograniczenia są łagodzone przez takie czynniki, jak m.in.: skupienie się na możliwie dobrym UX serwisu danej marki, dzięki czemu klient chętnie przegląda oferty, a później kupuje produkt, rekomendacje na bazie preferencji tego klienta czy podobnych, porównywarki produktów; wspierające treści (*content*), *live chat* z konsultantem itp.

Cena (ang. *price*), choć dziś potrafi zmieniać się dynamicznie⁵ w zależności od tego, z jakim klientem mamy do czynienia, oraz od tego, kiedy widzi

³ Targetowanie – proces polegający na selekcji konsumentów przez reklamodawców w celu określenia docelowej grupy odbiorczej, która w największym stopniu jest zainteresowana reklamowanym produktem, por. <https://pl.wikipedia.org/wiki/Targetowanie>.

⁴ Por. <http://uxdesign.pl/architektura-pozytywnych-doswiadczen>.

⁵ Por. <http://www.businessinsider.com/how-amazon-adjusts-its-prices-2015-1?IR=T>.

on ofertę, została najmniej dotknięta przez zmiany technologiczne czy też sam Internet. Oczywiście, pojawienie się porównywarek produktów i cen czy ogólny wzrost konkurencji (dziś klient może zamówić dany produkt na wielu rynkach, także zagranicznych) w sposób istotny zaostrzyło grę rynkową i podniosło presję cenową. W efekcie pojawiły się na rynku nowe typy ofert cenowych, proponowanych obniżek cen, programy sprzedażowe, lojalnościowe itd. Wymusiło to także zwiększenie intensywności kontaktu z klientem i samej reklamy.

Nic więc dziwnego, że w efekcie wszystkich powyższych zmian potrzeby marek wymagały, aby ich producenci poszukiwali coraz to efektywniejszych metod dotarcia do klienta ze swoim przekazem. W świecie wielokanałowości (*multichannel*, *cross-channel*, *omnichannel*), gdzie każdy kanał to możliwość kontaktu z klientem, liczba przekazów, jakie otrzymują odbiorcy, istotnie rośnie. Obok tradycyjnych kanałów reklamy, takich jak telewizja, radio i prasa, doszły liczne kanały *online*, a także kolejne urządzenia, np. telefon komórkowy czy zegarek. W każdym z nich klient może natknąć się na typową reklamę, ale także na inny format targetowanego przekazu marketingowego. Nic więc dziwnego, że przyjmuje się, iż przeciętnie człowiek może widzieć nawet 5 000 treści reklamowych dziennie⁶.

Jest jednak także druga strona medalu. Klienci wraz z pojawianiem się Internetu i mediów społecznościowych uzyskali dostęp do potężnej platformy dzielenia się opiniami, komentowania, a więc także przekazywania informacji zwrotnej firmom. W ciągu 60 sekund na Facebooku pojawia się 3 mln wiadomości, blisko pół miliona tweetów, 280 tys. snapów itd.⁷ Oznacza to, że producenci marek zyskali niemal bezpośredni dostęp do ogromnej liczby informacji na temat oczekiwań klientów czy ich aktualnych ocen danego produktu, które udostępniają samodzielnie, a więc stanowią one szczerą deklarację na dany temat.

Sytuacja opisana powyżej ma dwie istotne konsekwencje. Po pierwsze, wymagają na firmach jeszcze lepsze targetowanie i jeszcze lepsze rozumienie potrzeb klientów. Co za tym idzie, potrzebne są narzędzia, aby w tej niewyobrażalnie wielkiej liczbie informacji zwrotnych szybko uzyskiwać potrzebne dane i umieć je odpowiednio przetworzyć. Po drugie, klienci narażeni na taką dużą liczbę komunikatów coraz chętniej korzystają z narzędzi typu Adblock.

⁶ Por. <http://cbi.hhcc.com/writing/the-myth-of-5000-ads>.

⁷ Por. <http://www.go-globe.com/blog/60-seconds>.

Adblock – deklaracja klientów

Adblock to rozszerzenia dla przeglądarek WWW *desktop* i *mobile*, które umożliwiają blokowanie wyświetlania reklam. Ich działanie polega na filtrowaniu treści reklamowych w postaci plików graficznych, animacji czy wyskakujących okien z wyświetlanych stron internetowych⁸. W 2016 roku 36% (7,6 mln) użytkowników Internetu w Polsce blokowało reklamy *online*. W roku 2015 było to 29% (5,9 mln użytkowników), a w 2014 roku – 25% (4,9 mln użytkowników). Najczęstszymi powodami skłaniającymi Polaków do korzystania z adblocków są agresywne kampanie reklamowe (78%)⁹.

Firmy potrzebują więc nie tylko narzędzi do efektywnej analizy potrzeb klientów, ich zachowań i odczuć, do komunikowania się z nimi oraz do zarządzania ich relacjami, ale także takich, które pozwolą ograniczyć poziom niezadowolonych użytkowników z prezentowanych reklam.

Technologia w służbie marketingu

Wszystkie powyżej opisane zmiany rynkowe i wyzwania wynikające z oczekiwań klientów stawiały przed markami nowe oczekiwania, takie jak np. konieczność bieżącej analizy informacji płynących od klientów w kontekście wszystkich aspektów działalności marki, konieczność analizy i szybkiego reagowania na oczekiwania klientów, w tym pozostawanie z nimi w stałym, szybkim kontakcie i wreszcie konieczność efektywniejszego targetowania przekazu reklamowego, a także oceny skuteczności tej komunikacji. Ze względu na wspomnianą wcześniej liczbę informacji, jaką dzielą się klienci między sobą każdego dnia, oraz specyfikę Internetu, gdzie informacje przesyłane są niemal w czasie rzeczywistym, a ewentualne kryzysy potrafią rozprzestrzeniać się na globalną skalę błyskawicznie, nikogo nie dziwi, że na rynku pojawiła się potrzeba narzędzi, które pozwoliłyby na równie szybką analizę zachowań klientów w sieci.

Oprócz prowadzenia analiz zachowań klientów w oparciu o posiadane przez każdą firmę bazy danych (zawierające informacje m.in. o dokonanych zakupach, odwiedzanych stronach czy przeczytanych mailach) firmy coraz częściej korzystają z narzędzi pozwalających śledzić opinie klientów wyrażane w przestrzeni wirtualnej, w tym w szczególności w mediach społecznościowych. Na rynku funkcjonuje niezliczona liczba narzędzi, których marki używają do wyszukiwania danych archiwalnych także w czasie rzeczywistym (*Real Time Search*),

⁸ Por. <http://iab.org.pl/adblocki/o-adblockach>.

⁹ Por. *Raport IAB: Zjawisko blokowania reklam*.

m.in. *Google Alerts*, *Google Trends*, *Google Insights*, *Social Mention*, *Addictomatic*, *Klout*, *NutShellMail*, *Social Pointer*, *HowSocable*, *Uptrends*, a na polskim rynku *Sotrender*, *Brand24*, *Newspoint* itd. Część z tych narzędzi pozwala na koncentrowanie się na analizie danych płynących jedynie z kanałów należących do typu *social media*, część pozwala jednak na obserwowanie wszystkich publicznie dostępnych informacji. Dzięki nim marki uzyskują m.in. wiedzę dotyczącą np. natężenia dyskusji na dany temat, statystyk dających pełniejsze zrozumienie tego, co dzieje się z monitorowanym produktem czy hasłem, identyfikują źródła największego zainteresowania oraz pozwalają na automatyczną analizę tzw. sentymentu wypowiedzi (pozytywna, neutralna, negatywna).

Wprowadzenie do analizy sentymentu

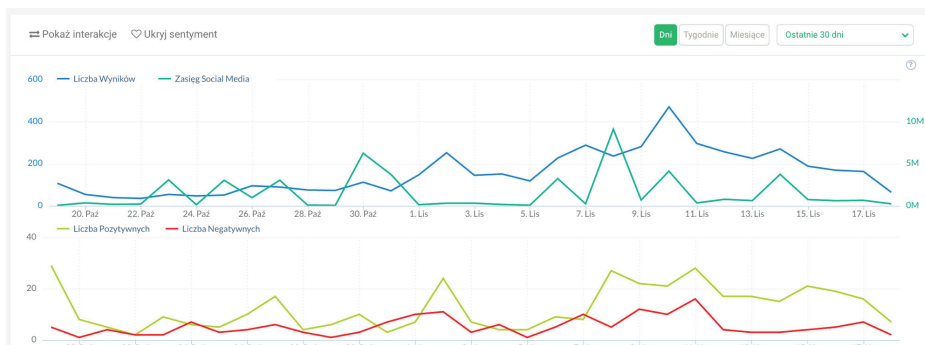
Ze źródeł wymienionych w poprzednich paragrafach producenci produktów konkretnej marki dowiadują się więc, że klienci omawiają jakiś temat (a nawet tego, co dokładnie piszą o danym produkcie lub usłudze), a dzięki analizie sentymentu zostaje określone nastawienie użytkownika sieci do omawianego tematu.

Analiza sentymentu odnosi się do automatycznych i półautomatycznych metod analizy tekstów. Podstawowe cele tych metod to identyfikacja i klasyfikacja fragmentów lub całych wypowiedzi ze względu na pojawiające się w nich słowa nacechowane emocjonalnie (Tomanek 2014). Analiza sentymentu opiera się bowiem na dwóch założeniach. Po pierwsze, niektóre wypowiedziane słowa wyrażają emocje. Po drugie, istnieją słowa, których wypowiedzenie może wywołać emocje (Pang, Lee 2008). Analiza sentymentu więc to analiza wydzwieku sentymentu opinii, polegająca na identyfikacji ładunku emocjonalnego wypowiedzi i zaszeregowaniu jej z tego punktu widzenia do jednej z trzech kategorii wypowiedzi: pozytywnej, negatywnej lub neutralnej. W rzeczywistości biznesowej analiza sentymentu wykorzystywana jest m.in. do przewidywania wyników sprzedaży (Liu 2007), a także do pozyskiwania opinii używanych później do oceny produktów oraz kupujących (McGlahen, Glance and Reiter 2010). Metody komputerowej analizy tekstu służą także coraz większej liczbie naukowców. Powszechnie wykorzystuje się je w analizach zjawisk kultury (Michel i in. 2011), lingwistyce (Lieberman i in. 2007), historii (Pagel, Atkinson, Meade 2007) czy antropologii (DeWall i in. 2011).

Wśród metod segregowania tekstów i wypowiedzi przyjmuje się dwie klasyfikacje: słownikową i statystyczną. Metoda słownikowa polega na logicznej analizie słów i fraz, uwzględniającej reguły leksykalne i składniowe rządzące danym językiem. W tym podejściu niezbędna jest znajomość: gramatyki

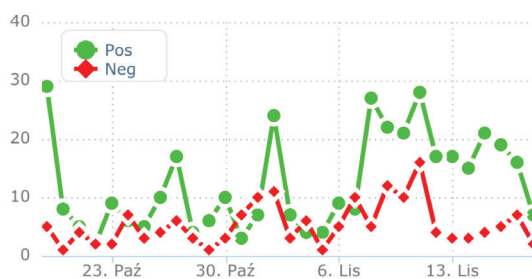
analizowanego języka, znaczeń, w jakich są używane występujące w nim słowa, specyfiki wypowiedzi związanej ze stosowanym słownictwem. W klasyfikacji słownikowej wykorzystywane są równoległe klasyfikacje wypowiedzi ze względu na występowanie w nich słów i wyrażań kluczowych; spełniające warunki logiczne dotyczące zawartości tekstu; wykorzystujące reguły leksykalne. Stosując metodę statystyczną, wykorzystuje się techniki ilościowe i traktuje tekst jako obiekt charakteryzowany przez dane ilościowe. Tekst jest więc reprezentowany jako wektor w wielowymiarowej przestrzeni wyznaczonej przez zbiór cech opisujących dokumenty. W tej metodzie wyróżnia się dwa podejścia: stosujące metody klasyfikacji opartej na idei uczenia się z nauczycielem (*supervised learning*) oraz bez nauczyciela (*unsupervised learning*) (Tomanek 2014).

W praktyce producenci danej marki wykorzystując narzędzia do monitorowania Internetu, otrzymują informacje na temat sentymentu poszczególnych wypowiedzi w 3 kategoriach: wypowiedź o wydźwięku pozytywnym, wypowiedź o wydźwięku negatywnym i wypowiedź o wydźwięku neutralnym. Z perspektywy firm cenne są zarówno informacje zagregowane – pokazujące ogólny trend i nastawienie wszystkich użytkowników komentujących dany temat, jak również informacje pochodzące od pojedynczego klienta. Te pierwsze pozwalają określić ogólną sytuację i percepcję firmy, marki lub produktu. Dla przykładu przewaga wypowiedzi o zabarwieniu negatywnym może sygnalizować narastający kryzys lub po prostu złą opinię dotyczącą omawianego tematu. Jeśli firma zechce rozwiązać problem, może analizując poszczególne pojedyncze wypowiedzi, określić, jakie jest źródło problemu, i odpowiednio na nie zareagować. Identyfikacja pojedynczej informacji o negatywnym zabarwieniu, nawet odosobnionej, pozwala producentowi danej marki zareagować na niezadowolony klienta w bardzo krótkim czasie (szczególnie jeśli firma prowadzi analizę w czasie rzeczywistym) czy też zahamować ewentualny kryzys, jeśli pierwsza wypowiedź będzie sygnałem większego problemu, który może dotknąć szerszej grupy odbiorców. W przypadku pozytywnych zgrupowanych wypowiedzi producent określonej marki może poszukać ambasadorów (osób, których wypowiedzi pozyskały największy zasięg, uruchomiły największą aktywność), a tym samym zdiagnozować najefektywniejsze kanały dotarcia ze swoim przekazem (*promotion*) do targetowanej grupy. Identyfikacja pojedynczej, pozytywnie nacechowanej wypowiedzi pozwala mu natomiast na równie pozytywną reakcję (podziękowanie, miły komentarz itd.), a co za tym idzie, pozyskanie zadowolonego (docenionego) klienta.



Rys. 1. Przykład wykresu zagregowanego – w górnej części liczba wyników i zasięg social media, w dolnej – wykres analizy sentymentu z liczbą pozytywnych i negatywnych wypowiedzi. Brand24.

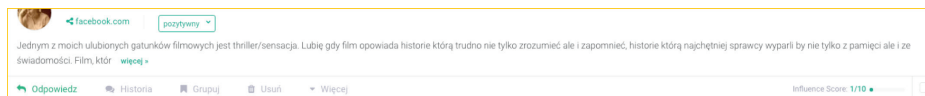
Sentyment



Rys. 2. Przykład innej prezentacji danych analizy sentymentu. Brand24.

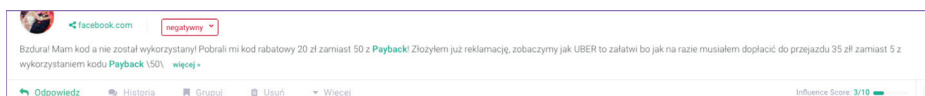
Poniżej podano przykłady pojedynczych wypowiedzi, zdefiniowanych przez narzędzie jako wypowiedź:

a) pozytywna¹⁰



¹⁰ Tekst przytoczonej z Facebooka wypowiedzi pozytywnej: *Jednym z moich ulubionych gatunków filmowych jest thriller/sensacja. Lubie gdy film opowiada historie którą trudno nie tylko zrozumieć ale i zapamiętać, historie którą najchętniej sprawcy wyparli by nie tylko z pamięci ale i ze świadomości. (...) [pisownia oryginalna – B.Ż.]*

b) negatywna¹¹



c) neutralna¹²



Rys. 3. Przykłady pojedynczych wypowiedzi użytkowników. Brand 24.

Analiza sentymentu – wyzwania

Jak zauważa wielu badaczy i praktyków, analiza sentymentu w obecnej formie, z wykorzystaniem narzędzi automatycznie kwalifikujących daną wypowiedź jako pozytywną, negatywną lub neutralną, nie jest idealna. Jej zwolennicy podkreślają, że korzystanie z technologii ułatwia szybką analizę dużej liczby danych w odróżnieniu od analizy przeprowadzanej przez człowieka. Pozwala również na ich agregację oraz bardzo czytelną prezentację. Dla przeciwników istotnym problemem jest fakt, że narzędzia te często się mylą, błędnie klasyfikując wypowiedzi.

Poniżej podano kilka przykładów błędnie zaklasyfikowanych wypowiedzi:

a) wypowiedź zaklasyfikowana pozytywnie, chociaż jej wydźwięk jest negatywny¹³



¹¹ Tekst przytoczonej z Facebooka wypowiedzi negatywnej: *Bzdura! Mam kod a nie został wykorzystany! Pobrali mi kod rabatowy 20 zł zamiast 50 z Payback! Złożyłem już reklamację, zobaczmy jak UBER to załatwi bo jak na razie musiałem dopłacić do przejazdu 35 zł! zamiast 5 z wykorzystaniem kodu Payback \50* [pisownia oryginalna – B.Ż.].

¹² Tekst przytoczonej z Facebooka wypowiedzi neutralnej: *Jeśli tankując na BP podam kasjerowi kupon z rabatem i aktywuje ekupon na Waszej stronie to nalicza mi się oba kupony? Chodzi o kupony ze zdjęć* [pisownia oryginalna – B.Ż.].

¹³ Tekst przytoczonej z Facebooka wypowiedzi zaklasyfikowanej pozytywnie, chociaż jej wydźwięk jest negatywny: *zajrzałam, bo myślałam o smartfonie, a tu --niemiła niespodzianka: wszystkie telefony spoza orange zniknęły, a te fundowane przez Orange mają simlocka, jak się od Państwa dowiedziałam. Trochę to jest nie fair...* [pisownia oryginalna – B.Ż.].

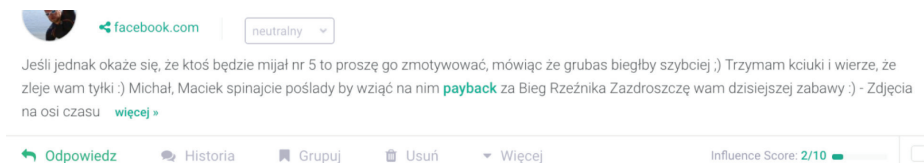
b) wypowiedź zaklasyfikowana negatywnie, chociaż jej wydźwięk jest pozytywny¹⁴



c) wypowiedź zaklasyfikowana neutralnie, chociaż jej wydźwięk jest negatywny¹⁵



d) wypowiedź zaklasyfikowana neutralnie, chociaż jej wydźwięk jest pozytywny¹⁶



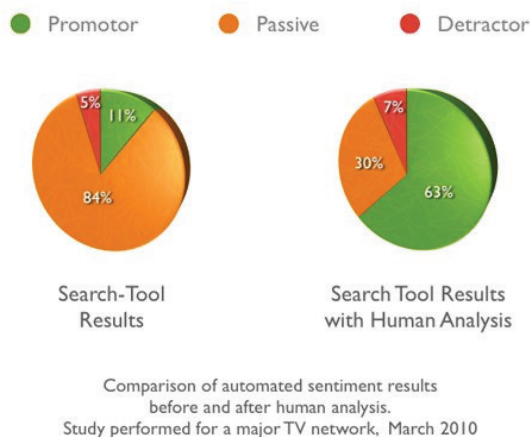
Rys. 4. Przykłady pojedynczych wypowiedzi użytkowników. Brand24.

Jak pokazuje rysunek 5, a także liczne badania przeprowadzane przez badaczy, analiza typu *supervised* (z nauczycielem) osiąga lepsze wyniki niż analiza typu *unsupervised* (bez nauczyciela). Skala problemu dla języka angielskiego różni się w zależności od badań, stopnia skomplikowania tekstu itd., jest jednak zauważalna (por. Bahrainin, Dengel 2015).

¹⁴ Tekst przytoczonej z Facebooka wypowiedzi zaklasyfikowanej negatywnie, chociaż jej wydźwięk jest pozytywny: *Uuuuuuuuuuu straszna, pomarańczowa PROMOCJA na produkty marki Lullalove. Tylko DZIŚ i tylko do PÓŁNOCY 10% rabatu przy zamówieniach za min. 50 PLN. Składając zamówienie wpisz KOD: LULLALOVE w polu Mam kupon rabatowy. Ofertą nie są objęte produkty już przecenione. #lullalove, #halloween, #promocja* [pisownia oryginalna – B.Ż.].

¹⁵ Tekst przytoczonej z Facebooka wypowiedzi zaklasyfikowanej neutralnie, chociaż jej wydźwięk jest negatywny: *Chce zrezygnować z programu payback i usunąć konto. Co muszę zrobić?* [pisownia oryginalna – B.Ż.].

¹⁶ Tekst przytoczonej z Facebooka wypowiedzi zaklasyfikowanej neutralnie, chociaż jej wydźwięk jest pozytywny: *Jeśli jednak okaże się, że ktoś będzie miał nr 5 to proszę go zmotywować, mówiąc że grubas biegłby szybciej ;) Trzymam kciuki i wierze, że zleje wam tyłki :) Michał, Maciek spinajcie pośladki by wziąć za nim payback za Bieg Rzeźnika Zazdroszczę wam dzisiejszej zabawy :) - Zdjęcia na osi czasu* [pisownia oryginalna – B.Ż.].



Rys. 5¹⁷.

Problem staje się tym większy, gdy mamy do czynienia z klasyfikacją wypowiedzi w języku potocznym czy też slangu. Dodatkowo trafność klasyfikacji spada w wypadku wypowiedzi w języku polskim, którego struktura gramatyczna istotnie różni się od struktury języka angielskiego. Poza tym, jak pisze Krzysztof Tomanek, „czym innym jest zliczenie słów wyrażających emocje, a czym innym rozumienie wypowiedzi przez pryzmat intencji autora wypowiedzi” (Tomanek 2014: 134).

Wnioski i rekomendacje

Trendem, który w tym roku stał się dominujący w tzw. *e-commerce* (handel elektroniczny), jest personalizacja. „Personalizacja będzie kluczem do sukcesu w *e-commerce* w 2017 roku. Musisz wiedzieć, czego klienci potrzebują, i to często zanim oni sami się tego dowiedzą. Jeśli rozpoznasz potrzebę i dostarczysz im rozwiązanie, zbudujesz rosnącą społeczność lojalnych odbiorców” – pisze brytyjska blogerka Kathryn Monkcom¹⁸.

W 2017 roku zakupy online będą wspierane przez analizę danych bardziej niż kiedykolwiek wcześniej dlatego, że dostosowanie oferty do kontekstu i okoliczności, w jakich znajduje się klient, staje się koniecznością, aby mu w ogóle

¹⁷ Rysunek pochodzi ze strony: <http://annamiotk.pl/analiza-sentymentu-a-bullshit-bingo/>.

¹⁸ Cytat pochodzi ze strony: <https://corporate.payu.com/blog/top-10-ecommerce-trends-2017>.

coś sprzedać – twierdzi John Komarek, założyciel agencji marketingu efektywnościowego Pixelter¹⁹.

Irina Weber, specjalistka od e-marketingu, radzi z kolei internetowym sprzedawcom, aby w kwestii personalizacji (i nie tylko) naśladowali Amazona i priorytetem uczynili wysokiej jakości obsługę klienta na każdym etapie procesu zakupowego²⁰.

Biorąc powyższe pod uwagę, nie ma wątpliwości, że znaczenie narzędzi do monitorowania Internetu będzie rosło podobnie jak znaczenie precyzyjniejszej umiejętności analizy sentymentu. Niezbędna wydaje się więc stała współpraca językoznawców i twórców takich narzędzi, by poziom trafności tych ostatnich istotnie wzrósł.

Pierwsze narzędzia do monitorowania Internetu pojawiły się w krajach anglojęzycznych, dlatego algorytmy, na których opierało się ich działanie, były dostosowane do specyfiki języka angielskiego, który prócz tego, że różni się od polskiego, także zmienił się przez ostatnie tysiąc lat bez porównania bardziej niż polszczyzna. Języki ewoluują w bardzo różnym tempie. Nasz ewoluuje stosunkowo powoli. Inne języki z naszego otoczenia zmieniają się szybciej²¹.

Język polski jako fleksyjny ma odmianę przez przypadki, występują w nim poza tym czasowniki dokonane i niedokonane, części mowy odmienne i nieodmienne, a do tego, jak mówi prof. Włodzimierz Gruszczyński, językoznawca ze Szkoły Wyższej Psychologii Społecznej, powołując się na literaturę przedmiotu, 5 rodzajów (a nie 3): męski, żeński, nijaki i jeszcze męskorzeczowy, męskożywotny, męskoosobowy²². Co więcej, język polski w odróżnieniu od angielskiego nie ma ustabilizowanego szyku zdania. Podczas gdy w języku angielskim obowiązuje schemat *podmiot – orzeczenie – dopełnienie* i jest on dość sztywny, w języku polskim podstawą jest, co prawda, struktura *podmiot – orzeczenie – dopełnienie*, ale istnieje duża dowolność co do kolejności umieszczania jej składników w zdaniu.

To wszystko sprawia, że nie było możliwe przeniesienie rozwiązań zagranicznych na rynek polski. Konieczne było utworzenie dedykowanych algorytmów, które radziłyby sobie ze specyfiką języka polskiego. Jak pokazują powyższe przykłady analizy sentymentu, ciągle jest wiele do zrobienia.

Dr Michał Kosiński, obecnie psycholog na Uniwersytecie Stanforda, opracował metodę tworzenia profili psychologicznych na podstawie zachowań na

¹⁹ Cytat pochodzi ze strony: <http://mamstartup.pl/e-commerce/10413/10-najwazniejszych-trendow-ecommerce-na-swiecie-w-2017-roku>.

²⁰ Por. <https://www.payu.pl/blog/10-najwazniejszych-trendow-ecommerce-w-2017-roku>.

²¹ Por. informacje zawarte na stronie: <http://www.newsweek.pl/polska/polski-najtrudniejszy-jezyk-swiata-jezykoznawca-odpowiada-newsweek,artykuly,352233,1.html>.

²² Op. cit., a także Mańczak 1956 oraz Saloni 1976.

Facebooku. Twórcy narzędzia wykorzystującego metodę dr. Kosińskiego za swój największy sukces uważają to, że potrafili przewidzieć doprowadzenie do Brexitu oraz zwycięstwo Donalda Trumpa w wyborach prezydenckich w Stanach Zjednoczonych²³. W związku z tym wydaje się, że technologia nie ma ograniczeń, a więc także narzędzia monitoringu Internetu z czasem będą w stanie nauczyć się czytać również emocje wypowiedane w języku polskim. To, czego im trzeba, to językoznawcy potrafiący opracować odpowiednie algorytmy dla języka polskiego.

Bibliografia

- Bahrainin S. A., Dengel A., 2015, *Sentiment analysis of text by capturing underlying sentiment patterns*, Web Intelligence 13 (2015), s. 53-68, DOI 10.3233/WEB-150309.
- Blossom J., 2009, *Content nation: Surviving and thriving as social media changes our work, our lives, and our future*, Wiley Publishing, Inc., Indianapolis.
- Gherghita-Mihaila D., 2016, *How is Social Media Influencing the Way we Communicate?* Acta Universitatis IOS Press.
- Giddens A., 2004, *Socjologia*. Warszawa.
- <http://annamiotk.pl/analiza-sentymentu-a-bullshit-bingo/>
- <http://cbi.hhcc.com/writing/the-myth-of-5000-ads/>
- <http://iab.org.pl/adblocki/o-adblockach/>
- <http://uxdesign.pl/architektura-pozytywnych-doswiadczen/>
- <http://www.businessinsider.com/how-amazon-adjusts-its-prices-2015-1?IR=T>
- <http://www.go-globe.com/blog/60-seconds/>
- <http://www.newsweek.pl/polska/polski-najtrudniejszy-jezyk-swiata-jezykoznawca-odpowiada-newsweek,artykuly,352233,1.html>
- https://pl.wikipedia.org/wiki/Por%C3%B3wnywarka_cen
- <https://pl.wikipedia.org/wiki/Targetowanie>
- <https://www.payu.pl/blog/10-najwazniejszych-trendow-ecommerce-w-2017-roku>
- Jorij A. J., 2012, *Omni-Channel Best Practices. How Companies can survive in an Omni-Channel World*, Brussels, March 28th.
- Kotler P., 2000, *Marketing Management*, Millennium Edition, Custom Edition for University of Pheonix, Prentice Hall.
- Kotler Ph., Keller K.L., 2012, *Marketing*, Poznań.
- Mańczak W., 1956, *Ile rodzajów jest w polskim*, „Język Polski”, nr 36, s.116-121.
- Qualman E., 2012, *Socialnomics: How Social Media Transforms the Way We Live and Do Business*.
- Raport IAB: Zjawisko blokowania reklam (dane bibliograficzne).
- Saloni Z., 1976, *Kategoria rodzaju we współczesnym języku polskim* [w:] R. Laskowski (red.), *Kategorie gramatyczne grup imiennych we współczesnym języku polskim*, Wrocław.

²³ Informacje pochodzą ze stron: <http://www.tvn24.pl/magazyn-tvn24/polak-odkryl-polityczna-bombe-atomowa-to-koniec-demokracji-jaka-znamy,79,1628>; <https://cambridgeanalytica.org/news>.

- Tomanek K., 2014, *Analiza sentymentu – metoda analizy danych jakościowych. Przykład zastosowania oraz ewaluacja słownika RID i metody klasyfikacji Bayesa w analizie danych jakościowych*, „Przegląd Socjologii Jakościowej”, nr 2, s. 118-136.
- Zolty K., 2015, *Analiza sentymentu*, <https://prezi.com/saypub4fu3iz/analiza-sentymentu/>;
„Acta Universitatis Danubius. Communicatio”, 2016, vol. 10, nr 1, s. 74-83.

Language in the service of marketing: Impact of technology on sentiment analysis

Summary

The aim of the article is to discuss the usefulness and language-specific limitations of automatic sentiment-analysis tools being used by marketing departments, and to argue for regular collaboration between these tools' developers and linguists. Automatic tools to analyse sentiment have emerged with the rise of the Internet and its spread in sales, marketing and communication, but their continued expansion is contingent on linguistic lore and capability for automatic linguistic analysis.

Key words: marketing, language of advertisement, sentiment analysis

Słowa kluczowe: marketing, język reklamy, analiza sentymentu