

*LESZEK SZYMAŃSKI**

WYBRANE ASPEKTY BADAŃ KWANTYTATYWNYCH POGAWĘDEK SIECIOWYCH

SELECTED ASPECTS OF QUANTITATIVE STUDIES ON ONLINE CHATS

Abstract

The article summarizes research on Internet chat room lexis. The author postulates that chatters utilize non-standard spelling and that chat room lexis reveals certain characteristics. In his opinion, these may be measured with corpus linguistic methodology, which he presents in the text.

Keywords: chat, language corpus, lexis

Słowa kluczowe: czat, korpus językowy, leksyka

1. WPROWADZENIE

Internet to sieć połączonych ze sobą komputerów. Dzięki niej można porozumiewać się mimo barier, jakie stwarza dystans fizyczny dzielący uczestników aktu komunikacyjnego. W sieci tej powstały nowe, interesujące środowiska komunikacyjne, korzystające ze zróżnicowanych form interakcji za pośrednictwem komputerów. Wraz ze wzrostem popularności wśród użytkowników stały się one także przedmiotem badań naukowych.

Niniejszy tekst z założenia stanowi zapowiedź monografii poświęconej analizie korpusowej leksyki z czatu internetowego¹. Znajdują się w nim również nawiązania do wcześniejszych publikacji wprowadzających do omawianego zagadnienia [Szymański, 2009; 2011; 2012], w których przedstawiono szczątkowe wyniki przedmiotowych badań. Czytelnik artykułu w pierwszej kolejności zapozna się z metodologiami badań komunikacji w Internecie stosowanymi dotychczas oraz z metodologią wykorzystaną w referowanym studium; w dalszej części zostaną zaprezentowane wybrane fragmenty wykonanych analiz.

* Uniwersytet Zielonogórski, Instytut Neofilologii, Zakład Filologii Angielskiej.

¹ Artykuł powstał na podstawie wygłoszonego 15 listopada 2011 r. na konferencji w Łodzi referatu, w którym zapowiedziano ukazanie się monografii poświęconej leksyce czatu internetowego.

2. DOTYCHCZASOWA METODOLOGIA BADAŃ NAD KOMUNIKACJĄ W INTERNECIE

Dotychczasowe analizy językoznawcze komunikacji w Internecie prowadzono na trzy sposoby [Szymański, 2011]. Po pierwsze, ten rodzaj komunikacji ujmowano przede wszystkim w sposób holistyczny. Oznacza to, że komunikacja w Internecie jest traktowana jako jedna całość. Badacze nie wyodrębniają poszczególnych form komunikacji w Internecie ani zróżnicowania środowisk komunikacyjnych [Szymański, 2011, s. 179]. Drugą metodę dociekań językoznawczych w tym obszarze stanowi opis zmanierowanych właściwości tego typu interakcji. Doprowadziło to do powstania wielu list odstępstw od reguł normatywnych, zwłaszcza dotyczących zapisu. Jako trzecią metodę wymienia się traktowanie komunikacji w Internecie jako hybrydy mowy i pisma, szczególnie w odniesieniu do czatu. Językoznawcy i komunikolodzy doszukują się w niej cech obu tradycyjnych kanałów komunikacyjnych.

3. METODOLOGIA WYKORZYSTANA W REFEROWANYM BADANIU

W omawianym badaniu komunikacji na czacie internetowym wykorzystano metody językoznawstwa korpusowego [zob. Piotrowski, 2000]. W celu przeprowadzenia analiz utworzono korpus języka używanego w Internecie².

Wybór metodologii językoznawstwa korpusowego spowodowany był dążeniem do otrzymania wyników opartych na konkretnych i niepodważalnych danych. Korpusy dostarczają bowiem danych empirycznych, co sprawia, że jest to metodologia obiektywna. Ponadto uzyskane dane statystyczne dostarczają informacji na temat zjawisk typowych i powszechnych dla badanego języka [Biber, 1996; Sambor, 2001].

Referowane badanie zawężono do jednej z form komunikacji w Internecie, czatu, z kilku powodów. Po pierwsze, podejście holistyczne do komunikacji w Internecie dałoby nierzetelne wyniki ze względu na jej gatunkowe zróżnicowanie. Po drugie, uzyskano dostęp do unikatowego materiału językowego – obszernych, oryginalnych zapisów pogawędek na czacie internetowym. Po trzecie, nie zarejestrowano wcześniej istnienia takiego korpusu, stąd pomysł na jego utworzenie i przeprowadzenie badań językoznawczych.

Omawiane w tym tekście badanie dotyczyło trzech sfer komunikacji na czacie internetowym. Po pierwsze, przyjrano się zapisowi form językowych, uwzględniając ilościowe badania odstępstw graficznych od norm językowych³. Po drugie, na podstawie szczegółowych analiz elementów leksyki ustalono pewne tendencje

² Pierwszą oficjalną informację o tym korpusie podano podczas zjazdu Polskiego Towarzystwa Językoznawczego w Opolu [zob. Szymański, 2009].

³ Dokładniejszy opis tego badania znajduje się w publikacji: Szymański, 2012.

słowotwórcze oraz zachowania językowe internautów⁴. Po trzecie, zweryfikowano zjawisko domniemanej hybrydy mowy i pisma, na czym później oparto wstępną definicję pogawędki internetowej jako gatunku [Szymański, 2009].

W dalszej części artykułu znajdzie czytelnik nawiązania do dwóch pierwszych z wyżej wymienionych zagadnień.

4. MATERIAŁ BADAWCZY

Materiał badawczy stanowił autorski korpus polskojęzycznego czatu internetowego. Został on już opisany w dwóch wcześniejszych publikacjach [Szymański, 2009; 2011]. W dalszej części bieżącego tekstu znajdzie czytelnik wybrane, reprezentatywne dla całego projektu, fragmenty analiz. Mają one na celu przybliżenie założeń, rozmiaru i charakterystyki eksperymentu.

Korpus Warsztat (taką nazwę przyjęto od nazwy czatu, z którego pochodzą próbki języka [Szymański, 2009]), można nazwać materiałem unikatowym, różni się on bowiem od innych istniejących korpusów wypowiedzi z czatów internetowych. Po pierwsze, czat Warsztat nie był zdominowany tematycznie. Po drugie, rejestrowano spontaniczne wypowiedzi internautów. Archiwizowane pogawędki nie były pierwotnie przeznaczone do badań, przez co uniknięto wystąpienia tzw. „paradoksu obserwatora”. Po trzecie, zapis tych pogawędek był prowadzony w sposób ciągły. Nie wybierano żadnych szczególnych momentów czasowych, które mogłyby mieć wpływ na wypowiedzi użytkowników.

Korpus Warsztat tworzą zapisane pogawędki internetowe prowadzone w okresie między 20 lutego 2004 r., godz. 17:32 a 27 marca 2006 r., godz. 22:54. Program do analiz statystycznych tekstów wyliczył, że korpus Warsztat składa się z 1 629 823 wyrazów⁵.

5. ANALIZY KWANTYTATYWNE⁶

5.1. Wybrane informacje statystyczne korpusu

Utworzony korpus wypowiedzi z czatu internetowego zestawiono następnie z dostępnym korpusem wzorcowym, zawierającym próbki języka znormalizowanego. W tym celu posłużono się korpusem *Słownika frekwencyjnego polszczyzny współczesnej* [Kurcz i in., 1990]. Korpus wzorcowy stanowi z założenia zbiór

⁴ Omówione w publikacji: Szymański, 2011.

⁵ Więcej informacji na temat edycji materiału, miejsca korpusu Warsztat w typologii korpusów T. Piotrowskiego [2003], jak również charakterystykę użytkowników oraz samego czatu znajdzie czytelnik w dwóch publikacjach: Szymański, 2009; 2011 oraz w zapowiadanej monografii.

⁶ W tej części pracy zestawiono wybrane fragmenty analizy, które mogą być częściowo znane z wcześniejszych publikacji poświęconych omawianym zjawiskom [Szymański, 2009; 2011; 2012].

tekstów reprezentatywnych dla danego języka, w tym przypadku dla języka polskiego. Stąd też zestawienie porównawcze utworzonego korpusu pogawędek internetowych z korpusem wzorcowym pozwala wynotować różnice między stylami w obu zbiorach.

Korpus Warsztat składa się z trzykrotnie większej liczby wyrazów (1 629 823) niż korpus wzorcowy (526 818). Na podstawie analiz statystycznych wykazano, że korpus internetowy charakteryzuje większe bogactwo leksykalne niż korpus standaryzowany [Szymański, 2012].

5.2. Analiza wyrazów kluczowych

Dysponując korpusem zrównoważonym, wykonano także badanie stu wyrazów kluczowych. Takie wyrazy odznaczają się niezwykle wysoką frekwencją w badanym korpusie w porównaniu z korpusem wzorcowym, w związku z czym stanowią zbiór wyrazów typowych dla badanego tekstu⁷. Wyodrębniono następujące grupy wyrazów kluczowych [za: Szymański, 2012]:

- a) niewyrażające konwencjonalnych znaczeń w języku polskim (ornamentacja wypowiedzi bądź część emotikonu), np. *P* pochodzące od *:P* lub *D* od *:D*,
- b) zapisane bez znaków diakrytycznych, np. *kolo*, *bede*, *sie*, *moze*,
- c) imiona własne użytkowników, np. *krzysio*, *mariusz*, *Nirgal*,
- d) wyrazy obcojęzyczne, np. *the*, *nick*,
- e) formy skrócone, np. *www*, *bry*,
- f) onomatopeje, np. *hmmm*, *hehehe*.

Z powyższego zestawienia klas wyrazów kluczowych można wywnioskować, że w wyrazach charakterystycznych dla komunikacji na czacie internetowym dają się zaobserwować pewne właściwości. Po pierwsze, mamy tu do czynienia z próbą wyrażania emocji w realizacji graficznej, którą uczestnicy czatu mają do dyspozycji. Dowodem występowania tego zjawiska jest obecność zarówno emotikonów, jak i onomatopei. Nadto internauci dążą do pośpiechu w komunikacji, wynotowano więc formy skrócone oraz wyrazy w pisowni bezdiakrytycznej. Ta druga powstaje ze względu na to, że czatującym nie chce się wcisnąć dodatkowego klawisza, by uzyskać znak diakrytyczny, spowalnia to bowiem, w ich mniemaniu, proces komunikacji. Warto też zaznaczyć, że mimo możliwej ambiwalencji znaczeń (np. *czesć* może oznaczać *czesć* lub *część*) uczestnicy czatu nie mają problemów z dezambiguacją znaczeń. W takich sytuacjach znaczącą rolę odgrywa kontekst [Szymański, 2012]. Pośpiech daje się także zaobserwować, gdy przyjrzymy się nienormatywnie zapisanym małą literą imionom własnym. Jest to kolejny dowód na pośpiech towarzyszący komunikacji internautów, którym nie chce się wcisnąć dodatkowego klawisza. Należy także zaznaczyć, że obecność grafemów wyrażających emocje

⁷ Definicje za M. Scottem [2004–2011].

oraz wyrazów dźwiękonaśladowczych świadczy o tym, że czat jest wykorzystywany jako substytut spontanicznej komunikacji bezpośredniej [Szymański, 2007–2008].

5.3. Analiza zapisu form językowych

Wykorzystując utworzony korpus, zanalizowano także nienormatywny zapis form językowych. Badaniu poddano właściwości charakterystyczne dla pisemnej komunikacji w Internecie, takie jak stosowanie znaków diakrytycznych, wielkich i małych liter oraz fonetyzacja zapisu [Szymański, 2012].

Internauci zwyczajowo piszą małymi literami. Potwierdza to choćby zapisywanie w ten sposób nazw zwyczajowo pisanych wielkimi literami, np. *warszawa*, *krzysio*, *polska*. W żargonie czatu pisownia wielkimi literami stanowi substytucję podniesionego głosu lub krzyku [Szymański, 2007–2008]. Zdarzają się jednak wypowiedzi, w których wielkie litery nie mają za zadanie wyrażania tej funkcji, np. (z 12.10.2004, 23:34):

<#[IP][psv]hatchet> no to musisz byc BARDZO madry :)⁸

W cytowanym fragmencie nadawca za pomocą wielkich liter chce zwrócić uwagę odbiorcy na swoistą istotność konkretnego wyrazu w ciągu całej wypowiedzi, wyróżnić go spośród innych wyrazów zapisanych małymi literami. Można zatem wnioskować, że wielkie litery, zwłaszcza gdy służą do zapisu całej formy wyrazu, są wykorzystywane do oznaczania emfazy.

Analizy zapisu form językowych pozwalają na stwierdzenie, że występujące w Internecie odstępstwa od pisowni normatywnej są zwykle działaniem celowym. Służą one bowiem internautom do prób wyrażania elementów komunikacji ustnej w formie graficznej [zob. Szymański, 2007–2008]. Dzięki konwencjom charakterystycznym dla tej formy interakcji, np. reduplikacji znaków, nienormatywnemu stosowaniu wielkich liter czy pisowni fonetycznej, uczestnicy czatu mogą wyrazić to, co chcą, korzystając w tym celu wyłącznie z graficznej formy przekazu. Z kolei pisownia bez znaków diakrytycznych jest wynikiem niechlujstwa i wygodnictwa użytkowników, dla których ekspresja oraz szybkość komunikacji stają się ważniejsze niż dbałość o poprawny zapis.

W badanych próbkach języka występuje również zjawisko zabawy formą graficzną wyrazów. Można je nazwać *fonetyzacją zapisu* [Szymański, 2012], gdyż głównym zadaniem tego rodzaju pisowni nienormatywnej jest graficzne odwzorowanie mowy. W przywołanej publikacji jako ten rodzaj niekonwencjonalności potraktowano zabawową substytucję niektórych znaków (np. *sh* na *sz* = *szit* [*shit*], *q* na *k* lub *ku* = *wqurza* [*wkurza*]), spolszczony zapis wyrazów obcojęzycznych (np.

⁸ Zapis oryginalny z materiałów korpusu Warsztat. Celowo nie użyto tutaj kursywy do oznaczenia cytatu, gdyż istnieje prawdopodobieństwo, że bywa ona wykorzystywana w komunikacji pisanej w Internecie jako oznaczanie emfazy.

lał, fenkju, senkju, danke szyn) oraz reduplikację liter (np. *coooooooooo*). W tej samej publikacji zauważa się jednak pewne nieścisłości, szczególnie w przypadku zwielokrotnienia liter. Zwrócono bowiem uwagę na sytuacje, gdy graficznej reduplikacji ulegają nie tylko zwyczajowo wydłużane w mowie samogłoski akcentowane, lecz także samogłoski nieakcentowane (np. *cebulaaaaaaaaaaaaaa*), a nawet spółgłoski (np. *psychollllllllllll*) [Szymański, 2012].

Na podstawie materiału językowego z czatu Warsztat można poczynić pewne spostrzeżenia. Stwierdzono np. [Szymański, 2012], że pisownia nienormatywna stała się pewnego rodzaju wyróżnikiem komunikacji w Internecie, zwłaszcza na czatach. Wytworzenie własnych konwencji zapisu może być swoistą próbą manifestowania swojej odrębności i indywidualności bądź też wyróżnikiem przynależności grupowej uczestników czatu.

5.4. Analiza leksyki czatu

Oprócz analiz omówionych powyżej wykonano także badanie frekwencyjne leksyki czatu internetowego z wykorzystaniem korpusu Warsztat. W eksperymencie tym zwrócono uwagę na wyrażenia wykorzystywane w łączności fatycznej oraz kojarzone z nieoficjalnością. Poddano analizie [Szymański, 2011] 420 wyrazów w funkcji powitania, 299 form wyrazów w funkcji zwrotów pożegnalnych, 2291 wyrazów w funkcji podziękowań, 1097 wyrażen przeproszenia oraz 759 wyrazów uznanych za wulgarne, przywołano też 125 form obcojęzycznych oraz 67 form skróconych, które wypisano z pierwszego tysiąca wyrazów o najwyższej frekwencji. Ponadto wykonano badanie 7456 unikatowych pseudonimów użytkowników, częściowo sygnalizowane wcześniej [Szymański, 2009]⁹.

Wynotowane wyrazy (bądź wyrażenia) poddano badaniom ilościowym, jak również konkordancjom, czyli badaniom kontekstów, w których dane wyrazy występują w analizowanym korpusie. Uzyskane wyniki pozwoliły na ustalenie szeregu faktów, które są wymienione w dalszej części tekstu [za: Szymański, 2011].

Przede wszystkim obserwuje się usilne dążenie uczestników pogawędek w Internecie do nieoficjalności. Internauci manipulują wyrażeniami, w tym także oficjalnymi, i na różne sposoby zmniejszają ich oficjalność, np. poprzez złożone procesy słowotwórcze, mające na celu tworzenie zdrobnień, por.: *dziendoberek, witka, siemanko, dziekowka, nareczka*. W ten sposób nie stwarza się dystansu między nadawcą i odbiorcą (lub odbiorcami). Co więcej, powitanie pozostałych uczestników czatu wyrażaniem kojarzonym z nieoficjalnością wytwarza taką atmosferę już od początku aktu komunikacyjnego. W ten sposób nie tylko realizuje się funkcja fatyczna czy ludyczna, lecz także osiągnię jest cel pozajęzykowy, którym może być dążenie do uzyskania przyjaznej atmosfery.

⁹ Planowany jest oddzielny artykuł, w którym przedstawione zostaną wyniki analizy pseudonimów internautów, znanych także jako nicki.

Wynotowano również wyrazy, które powstały w wyniku zabawy językiem. Internauci wykorzystują w nich graficzność kanału przekazu. Można tu wymienić, np. *senkju, pseplasm, sorqqqqqqqq, pojaFFiam, tesh, jush*. Ich obecność w badanym korpusie zdaje się potwierdzać tezę, że internauci mają motywację dążenia do nieoficjalności i dowcipu.

W wyniku pośpiechu, który towarzyszy internautom, powstają formy przypadkowe, zawierające znaki intruzywne, np. *dziendobrt, cfześć, branocjka, dzwieki*. Obecność tych znaków jest przejawem braku dbałości o formę, gdy szybkość przekazu staje się ważniejsza niż poprawność językowa czy szacunek wobec adresata.

Interesujących spostrzeżeń dokonano w związku z obecnością wyrazów obcojęzycznych w badanym korpusie [Szymański, 2011]. Zaobserwowano, że wyrazy takie są stosowane przez Polaków w kontekstach polskojęzycznych w realizacji aktów mowy, np. *zdrastwujcie, bye, sayonara, thanks, spasiba, sorry*. O ile wtrącenia obcojęzyczne w powitaniach, pożegnaniach czy podziękowaniach mogą być motywowane dążeniem internautów do dowcipu, o tyle w funkcji przeproszenia takie wyrażenie może mieć inny cel. Dopatrzone są w tym wypadku trudności w przyznaniu się do winy i wyrażeniu skruchy. W związku z tym internauci uciekają się do zastosowania wyrażenia obcojęzycznego, z którym nie czują więzi emocjonalnej [Szymański, 2011]. Wyraz obcojęzyczny działa wówczas jako swego rodzaju tarcza ochronna przed niezamierzonym utraceniem własnego wizerunku w oczach odbiorcy. Ponadto odnotowano także przypadki stosowania wyrazów i zwrotów obcojęzycznych jako substytutów wyrazów i zwrotów polskich. Internauci używali wyrazów i zwrotów obcojęzycznych, aby popisać się znajomością języków obcych. Fakt celowego mieszania języków, również dla zabawy, wpływał w pewnym stopniu na nieoficjalność pogawędek.

W komunikacji na czacie pojawiają się również neosemantyzmy. Powstają one w wyniku podobieństwa fonetycznego do innych wyrazów, np. *nerka* w funkcji pożegnania jako wyraz podobny do pożegnania *narka*, bądź w wyniku derywacji, np. pożegnanie *nocka* od *dobranoc*. Występują także wyrażenia konwencjonalnie używane w mowie, np. *do zobaczenia*. Oczywiście jest, że w komunikacji tekstowej na czacie internetowym fizyczne zobaczenie kogoś jest w zasadzie niemożliwe, chyba że internauci żegnali się za pomocą tego zwrotu, mając na myśli zobaczenie się w świecie rzeczywistym. Takie przeniesienie zwyczaju może świadczyć o tym, że internauci traktują pogawędki na czacie jako substytut komunikacji mówionej bezpośredniej. Zauważono także [Szymański, 2011] tendencję do nominalizacji w realizacji aktów mowy. Internauci tworzą bowiem nazwy rzeczownikowe do nazywania czynności, np. *dziekowka, sorencje, witka*. Wygląda więc, jakby użytkownicy czatu czuli potrzebę przekazywania komunikatów za pomocą rzeczowników.

Uwagę zwraca także wykorzystanie skróconego zapisu wyrazów [Szymański, 2011]. Można wyróżnić dwa rodzaje wyrazów skróconych. Jednym z nich są znane

powszechnie formy skrócone, np. *vip*, *http*, *www*, *lol*, *brb*, *thx*. Są one używane i znane użytkownikom języka nie tylko na czatach internetowych, lecz również w innych dziedzinach życia. Drugim typem wyrazów skróconych są skróty lokalne. Zasięg ich użycia ogranicza się wyłącznie do konkretnej grupy użytkowników konkretnego czatu, np. *vold*, *hust*, *def*. Badania frekwencyjne występowania form skróconych dowiodły, że takie wyrazy nie są częste; w badanym korpusie Warsztat obecność form skróconych wyrażają wartości promilowe [Szymański, 2011].

W dalszej części referowanego badania, korzystając z definicji Macieja Grochowskiego [2000], przyjrano się kulturze wypowiedzi internautów. Ściślej mówiąc, zanalizowano wulgaryzmy występujące w badanym korpusie [Szymański, 2009; 2011]. Zebrany materiał umożliwił ustalenie stopnia wulgarności wypowiedzi na czacie internetowym. Wykazano, że wulgaryzmy w pogawędkach internetowych pełnią te same funkcje co wulgaryzmy w mowie [zob. Kowalikowa, 2000]. Jednakże w przypadku badanego materiału można mówić o względnej kulturze wypowiedzi [Szymański, 2009]. Na podstawie analizy wynotowanych wulgaryzmów pokazano, że nie są one liczne. Wynotowano 759 wyrazów uznanych za wulgarne (lub derywowanych od wulgaryzmów), co stanowi 0,54% wszystkich wyrazów z listy frekwencyjnej [Szymański, 2011]. Najczęstsze zaś wulgaryzmy nie wyrażają silnej wulgarności. Oznacza to, że chociaż nie są traktowane jako ordynarne, mogą być postrzegane jako nieeleganckie [Szymański, 2011]. Wśród tysiąca najczęstszych wyrazów znalazły się cztery wulgaryzmy (a właściwie przekleństwa): *kurde*, *kurna*, *cholera*, *qrde*.

Zbadano również nazwy własne internautów, czyli nicki. Są one nieoficjalnymi nazwami przybieranymi przez uczestników pogawędek. Stają się przez to identyfikatorami osobowymi określonych użytkowników; bywa też, że są przenoszone do świata rzeczywistego, gdzie konkretne osoby są identyfikowane pod nazwami przeniesionymi ze świata wirtualnego. Badanie tych nazw własnych miało na celu przedstawienie tendencji w ich tworzeniu oraz ustalenie, jakie informacje można z nich uzyskać na temat danego użytkownika. Na podstawie analizy wynotowanych 7456 unikatowych pseudonimów internautów ustalono [za: Szymański, 2009], że znajdują się wśród nich zarówno pojedyncze wyrazy, jak i nazwy kilkuwyrazowe. Znaczna ich część została zapisana nienormatywnie, tj. małą literą¹⁰. Wielkie litery w nickach pełniły raczej funkcję ornamentacyjną, np. <PiTeRaS>, <BuGGy>, <[PL]K4c2m4r>. Z kolei wykorzystanie cyfr w budowie pseudonimu ma głównie na celu identyfikację wieku użytkownika, np. <mody-85> (gdzie podano rok urodzenia) lub <MiLaK_18> (gdzie podano wiek liczony w latach). Internauci zawierają w nickach także informacje na temat swojego pochodzenia, zarówno narodowo-

¹⁰ Wg normy ortograficznej [Polański, 2006], nazwy osobowe, imiona własne powinny być zapisane wielką literą.

ściowego, np. <(PL)Johrel>, jak i miejscowego¹¹, np. <Ziomek-WLKP>. Ponadto, użytkownicy Warsztatu wyrażali w nickach hierarchię czatu, np. <![VIP]ManieK>.

6. ZAKOŃCZENIE

Na zakończenie warto podkreślić, że przytoczone wyżej analizy pogawędek internetowych zostały przeprowadzone z wykorzystaniem metodologii językoznawstwa korpusowego. Dzięki zastosowaniu korpusu dokonano oglądu rzeczywistych zachowań językowych Polaków, utrwalonych w tekstach internetowych pogawędek. Ponadto dane empiryczne umożliwiły także ustalenie, co jest częste i typowe w tej formie komunikacji. Badanie przywołane w niniejszym tekście stanowi swoisty głos w dyskusji nad właściwościami języka w komunikacji internetowej.

BIBLIOGRAFIA

- BIBER, Douglas, 1996, Investigating Language Use through Corpus-based Analyses of Association Patterns, „International Journal of Corpus Linguistics”, 1 (2), Amsterdam, s. 171–197
- GROCHOWSKI Maciej, 2000, Słownik polskich przekleństw i wulgaryzmów, Wydawnictwo Naukowe PWN, Warszawa
- KOWALIKOWA Jadwiga, 2000, Wulgaryzmy we współczesnej polszczyźnie, [w:] Język trzeciego tysiąclecia, red. G. Szpila, Kraków, s. 121–132
- KURCZ Ida, LEWICKI Andrzej, SAMBOR Jadwiga, SZAFRAN Krzysztof, WORONCZAK Jerzy, 1990, Słownik frekwencyjny polszczyzny współczesnej, Instytut Języka Polskiego PAN, Kraków
- PIOTROWSKI Tadeusz, 2003, Językoznawstwo korpusowe: wprowadzenie do problematyki, [w:] Językoznawstwo polskie. Stan i perspektywy, red. S. Gajda, PAN, Uniwersytet Opolski, Warszawa–Opole, s. 143–154
- POLAŃSKI Edward, 2006, Zasady pisowni, [w:] Polszczyzna na co dzień, red. M. Bańko, Wydawnictwo Naukowe PWN, Warszawa
- SAMBOR Jadwiga, 2001, Język polski w świetle statystyki, [w:] Współczesny język polski, red. J. Bartmiński, Lublin, s. 503–526
- SCOTT Mike, 2004–2011, Introduction to Oxford WordSmith Tools (dostępny w Internecie) http://www.lexically.net/downloads/version5/HTML/?keywords_info.htm (5.11.2011)
- SZYMAŃSKI Leszek, 2007–2008, Substytuty elementów parawerbalnych i niewerbalnych w komunikacji tekstowej w Internecie, [w:] Media – kultura – komunikacja społeczna, Olsztyn, s. 148–157

¹¹ Uczestnicy badanych pogawędek oznaczali domniemane miejsce swojego pochodzenia, np. <marcin_poznan>, lub to, w którym w danej chwili przebywali, np. <Krzysio_dom>. Oczywiście takie informacje należy traktować jako przypuszczalne, gdyż nie sposób ich zweryfikować.

- SZYMAŃSKI Leszek, 2009, Analiza wypowiedzi z czatu internetowego, „Biuletyn Polskiego Towarzystwa Językoznawczego”, 65, s. 165–174
- SZYMAŃSKI Leszek, 2011, Empiryczna analiza leksyki czatu internetowego, „Język Polski”, 91, 2–3, s. 179–187
- SZYMAŃSKI Leszek, 2012, Konwencje zapisu wyrazów na czacie internetowym, „Język Polski”, 92, 1, s. 20–28

Leszek Szymański

WYBRANE ASPEKTY BADAŃ KWANTYTATYWNYCH
POGAWĘDEK SIECIOWYCH

Streszczenie

W artykule przedstawiono wybrane fragmenty badania leksyki czatu internetowego z wykorzystaniem metodologii korpusowej. Na początku autor przedstawia dotychczasową metodologię obecną w badaniach komunikacji w Internecie, po czym przechodzi do opisu metodologii wykorzystanej w referowanym badaniu. Czytelnik poznaje tutaj najistotniejsze cechy utworzonego korpusu pogawędek internetowych. Następnie autor przechodzi do opisu wybranych zagadnień kwantytatywnych badań leksyki czatu. Rozpoczyna od omówienia statystyk badanego korpusu. Dalej przedstawia analizę niestandardowego zapisu wyrazów, po czym przechodzi do zrelacjonowania ważniejszych obserwacji odnośnie charakterystycznych właściwości leksyki czatu.

SELECTED ASPECTS OF QUANTITATIVE STUDIES
ON ONLINE CHATS

Summary

The article discusses selected fragments of a corpus-based analysis of Internet chat room lexis. First, the author presents previously used methodologies for Internet communication studies. After that, he describes the methodology he used in his research. At this point, the reader becomes acquainted with the most crucial characteristics of the created corpus of Internet chats. Further, the author goes on to delineate selected elements of quantitative studies on chat room lexis. He begins with the portrayal of corpus statistics. Then, he proceeds to discuss non-standard spelling, after which, he reports on the most important observations as for the characteristics of chat room lexis.