



Polszczyzna w dobie cyfryzacji

OGÓLNOPOLSKA INTERDYSCYPLINARNA KONFERENCJA POŚWIĘCONA KOMUNIKACJI JĘZYKOWEJ
www.forumkultury.slowa.pl; e-mail: forumkultury.slowa@gmail.com

Organizatorzy



Miejsce obrad

Uniwersytet Jana Kochanowskiego
Instytut Filologii Polskiej, aula B1
ul. Świętokrzyska 15G
25-406 Kielce

REFERATY

Edwin Bendyk (Collegium Civitas), *Spoleczne i społecznościowe kształtowanie wiedzy*

Rozwój społecznościowych form wytwarzania, agregowania i upowszechniania wiedzy, którego spektakularnym przykładem jest Wikipedia doprowadził do paradoksalnych konsekwencji. Z jednej strony nowe formy współpracy opartej na wymianie komunikacyjnej w sieciach cyfrowych mediów umożliwiły ekspresję rzeszom zaangażowanych amatorów. Tym samym podważony został monopol „tradycyjnych” instytucji organizujących ruch wiedzy. Z drugiej jednak strony współpraca bez udziału formalnych hierarchii wymaga innych form stabilizacji i prowadzi do konserwatyizmu. W rezultacie społeczne i społecznościowe formy organizowania ruchu wiedzy wchodzą w dynamiczne relacje konfliktu i współpracy. W referacie skoncentruję się na tej właśnie paradoksalnej dynamice.

Andrzej Jacek Blikle (Rada Języka Polskiego), *Cyfrowy tezaurus polskich terminów informatycznych*

Tezaurus, o którym mowa w tytule, to bazodanowa aplikacja z dostępem przez Internet, która w zamierzeniu jej twórców ma posłużyć do ujednoczenia, uściślenia i upowszechnienia polskojęzycznej terminologii informatycznej. To zadanie wydaje się szczególnie ważne w epoce postępującej cyfryzacji życia na świecie, w tym i w Polsce. Tezaurus powinien stać się narzędziem nie tylko dla informatyków, ale dla wszystkich osób mających do czynienia z produktami, zjawiskami i pojęciami informatycznymi w obszarze nauki, edukacji, gospodarki a także stosowania i stanowienia prawa.

Iwona Burkacka (Uniwersytet Warszawski), *Język memów internetowych*

Memy internetowe są popularnym sposobem komentowania rzeczywistości, podkreślania i budowania wspólnotowości oraz formą ekspresji. Świetnie wpisują się w zjawiska zachodzące we współczesnej kulturze i sztuce. Związki z kulturą kontrmówienia i recyklingiem kulturowym oraz przejawy intertekstualności są widoczne w języku memów: warstwie językowej i graficznej, a także w relacjach zachodzących między słowem a obrazem. W memach występują także nowe elementy językowe (m.in. neologizmy słowotwórcze, neofrazeologizmy) i słownictwo środowiskowe oraz liczne przykłady modyfikacji przysłów, frazeologizmów, skrzydlatych słów czy cytatów.

Bartłomiej Chaciński („Polityka”), *Memy jako kanon kulturowy i ich wpływ na język*

Czym jest zjawisko memów internetowych dla języka? W jaki sposób formy rozprzestrzeniania się memów zmieniają język, którym posługujemy się w sieci i poza nią? Zjawisko to zostanie omówione na przykładzie konkretnych słów, które wyszły z memów, zmieniły się dzięki nim lub za ich pomocą zyskały większą niż dotąd popularność, a raczej: większy zasięg, bo to właśnie kluczowe pojęcie dzisiejszej sieci.

Mirosław Filiciak (Uniwersytet SWPS), *Język zmarginalizowany. Komunikacja w epoce kodu i algorytmów*

Wobec informacyjnego nadmiaru coraz większy wpływ na komunikowanie się w sieci odgrywają mechanizmy filtrujące, takie jak algorytmy serwisów społecznościowych. Stworzone na użytek marketingu i mechanizmów rekomendacyjnych, zamykają nas w „bańkach”, wyznaczając horyzont tego, z kim i jak prowadzimy dyskusje. W efekcie – jak przekonują badacze z kręgu materialistycznej teorii mediów – językiem pełniącym dziś często funkcję nadrzędną wobec języków naturalnych staje się kod komputerowy. Zyskuje on niemal „magiczną” moc performatywną, a równocześnie pozostaje nie tylko nietransparentny, ale też często w ogóle niezauważalny dla osób biorących udział w komunikacji.

Włodzimierz Gruszczyński (IJP PAN), *Korpusy narzędziem pracy historyka języka*

W referacie przedstawię istniejące i powstające obecnie korpusy tekstów polskich z różnych epok (przede wszystkim: korpus staropolski, polszczyzny XVI w., tekstów polskich XVII-XVIII w., korpus „gronowy” z XIX w., ale też korpusy inne, jak np. korpus średniowiecznych rot sądowych czy XVI-wiecznych tłumaczeń ewangelii). Omówię sposoby posługiwania się tymi korpusami (wyszukiwarki, język zapytań). Pokażę na przykładach możliwości zastosowania korpusów w badaniach historycznojęzykowych. Przedstawię także projekt integracji największych korpusów oraz narzędzia umożliwiające korzystanie jednocześnie z nich wszystkich (przede wszystkim Chronofleks).

Agata Hacia (Fundacja PRO.PL), *Co myślą o języku polskim, co czują i robią w związku z nim aktywni użytkownicy Internetu*

W referacie podejmę próbę charakterystyki postaw wobec języka prezentowanych przez aktywnych użytkowników Internetu, zrzeszających się w grupy oparte na przynależności do grupy zawodowej, na dzieleniu hobby, na wspólnocie poglądów. Odniosę się do klasycznie wyodrębnianych komponentów postaw (emocjonalnego, poznawczego, behawioralnego), opiszę proporcje, w jakich komponenty te są reprezentowane w analizowanym materiale, a także wyciągnę wnioski porównawcze.

Mariusz Herma („Polityka”), *Tłumaczenia automatyczne: jakość i ich wpływ na komunikację w sieci*

Serwis Google Translate odpowiada na miliardy zapytań dziennie, podobnie algorytmy językowe Facebooka, Twittera czy chińskiego WeChata. Kto i dlaczego najczęściej z nich korzysta? Jakiej jakości tłumaczenia oferują i jaki jest potencjał ich dalszego rozwoju?

Marta Juza (Uniwersytet Pedagogiczny w Krakowie), *Od języka internetu do pragmatyzmu komunikacji*

Gdy internet zaczynał upowszechniać się w życiu społecznym, na przełomie XX i XXI wieku, dostęp do niego był jeszcze dość rzadki, a liczba dostępnych za pośrednictwem sieci usług komunikacyjnych niewielka, zaczynał kształtować się odrębny socjolekt, specyficzny dla wszelkich form komunikacji internetowej. Był on odrębny dla różnych języków narodowych, choć można było w nim dostrzec wyraźny wpływ języka angielskiego. Obecnie, gdy w rozwiniętych krajach świata internetu używa większość ludzi, przestaje on być traktowany jako odrębna sfera rzeczywistości, a staje się częścią życia codziennego. Jest też postrzegany jako przede wszystkim narzędzie, które ma umożliwiać komunikację między ludźmi. Tym samym następuje też zwrot od odrębnego internetowego socjolektu, którego użytkownicy musieli się uczyć, w kierunku takiego stosowania języka, które umożliwi porozumienie przy jak najmniejszym nakładzie sił. W referacie zostanie omówiona ta ewolucja komunikacji językowej przy użyciu internetu, jak również obecne zjawiska charakterystyczne dla tej komunikacji, wraz z wyjaśnieniem ich przyczyn.

Ryszard Kluszczyński (Uniwersytet Łódzki), *Hipermedralność i remediacja druku*

Współczesne praktyki twórcze w coraz większym stopniu odwołują się do partycypacji odbiorczej, z jednej strony sięgając po kolektywnie tworzone bazy danych, z drugiej strony prowokując interaktywne zachowania odbiorcze. W rezultacie wytwory tych praktyk tracą stabilną, trwałą, ostateczną formę, która jest zastępowana przez procesy i wydarzenia. Druk jako ostoja trwałości musi się zmierzyć z wyzwaniem tymczasowości. Formy tekstowe są wypierane przez struktury hipertekstowe. Zarazem wytwory praktyk twórczych, obok niestabilności strukturalnej, charakteryzuje też hybrydyczność materiałowa. W rezultacie stają się one inter-, multi-, transmedialne. Hipertekstualność przybiera postać hipermedialności, a twórczość językowa staje się kreacją hipermedialną.

Katarzyna Kłosińska (Uniwersytet Warszawski), *Norma językowa w dobie cyfryzacji*

W referacie zostanie omówiony wpływ nowych mediów na funkcjonowanie normy językowej oraz zostanie przedstawiona próba zredefiniowania pojęcia normy językowej.

Joanna Kowalczyk (Uniwersytet Jana Kochanowskiego w Kielcach), *Komunikacja urzędowa w dobie cyfryzacji*

W wystąpieniu zostaną omówione główne tendencje w sposobie komunikowania pomiędzy urzędem a obywatelem, spowodowane powstaniem e-administracji. Przeniesienie części zadań publicznych do przestrzeni cyfrowej wpłynęło m.in. na eklektyzm języka oraz znaczące wyrównanie pozycji urzędu oraz obywatela w wielu wymiarach.

Piotr Kubiński (Uniwersytet Warszawski), *Ludyczny charakter używania języka w grach*

Wystąpienie podejmuje temat ludycznego użycia w grach wideo zarówno z perspektywy tłumacza, jak i twórcy jej scenariusza oraz dialogów. Celem referatu jest przybliżenie słuchaczom samego zjawiska, ukazanie jego specyfiki oraz zaprezentowanie gier wideo jako bogatego obszaru potencjalnych badań językoznawczych.

Iwona Loewe (Uniwersytet Śląski), *Język w mediach cyfrowych*

Internet nie generuje nowego języka etnicznego, ale jest nowym sposobem jego manifestacji – jak nigdy dotąd utrwala polszczyznę współczesną we wszystkich stylach funkcjonalnych. Cechy technologiczne sieci i mediów cyfrowych powodują natomiast, że pośrednio mogą one wpływać na pewne aktywności użytkowników języka. Chciałabym w wystąpieniu zwrócić uwagę na niektóre z nich.

Marek Łaziński (Uniwersytet Warszawski), *Słowa klucze polszczyzny doby cyfryzacji w statystyce i w świadomości społecznej*

Próby wyszukania i opisanie słów najważniejszych podejmowane były od początku cywilizacji. Co najmniej od czasów średniowiecznych – od konkordancji biblijnych – próbuje się także szukać takich słów metodami statystycznymi, które od końca XX wieku można stosować w ogromnych ogólnie dostępnych zasobach tekstowych – korpusach. W referacie przedstawione zostaną wyniki poszukiwań słów najczęstszych i najważniejszych współczesności w projektach UW „Słowa dnia” i „Słowa roku” w szerszej perspektywie badań kluczowości pojęć w lingwistyce kulturowej. Centralnym punktem rozważań będzie porównanie wyników kluczowości słów w badaniach statystycznych i ankietowych (w tym plebiscytowych).

Jacek Mazurczak, Jacek Purski (Instytut Bezpieczeństwa Społecznego), *Język w sieci jako narzędzie radykalizacji*

Radykalizacja to złożony proces, w którym osoba lub grupa osób zaczyna przyjmować skrajną, opartą na różnych formach przemocy, ideologię lub przekonania. To odrzucenie przyjętego przez ogół społeczeństwa status quo i negowanie podstawowych wartości demokratycznych oraz rosnąca skłonność do używania przemocy dla osiągnięcia celów ideologicznych. W ramach wystąpienia zaprezentujemy w jaki sposób język, w tym w szczególności mowa nienawiści, wykorzystywane są przez ruchy ekstremistyczne w ich działalności. Przedstawione zostaną najważniejsze organizacje i inicjatywy nieformalne o charakterze radykalnym działające w Polsce, a także kody jakimi się posługują.

Marlena Modzelewska, Kamila Węglarska (Narodowe Centrum Kultury), *Emocje w sieci. Analiza internetowych dyskusji związanych z rocznicą odzyskania niepodległości*

Od 2016 roku Narodowe Centrum Kultury bada sposoby obchodzenia Święta Niepodległości. W roku 2018 jednym z komponentów badań była analiza sieci społecznych (*Social Network Analysis*) przeprowadzona na podstawie danych pochodzących z lat 2017–2018. Przy wyszukiwaniu dyskusji posłużyliśmy się następującymi słowami kluczowymi oraz ich odmianami: *11 listopada, Święto Niepodległości, Marsz Niepodległości, niepodległa, niepodległość*. Dzięki tej analizie zidentyfikowaliśmy liderów dyskusji, dowiedzieliśmy się, gdzie internauci najczęściej komentowali kwestie związane ze Świętem Niepodległości oraz jak wydarzenia bieżące wpływały na dyskusję w sieci. W kolejnym kroku zrealizowaliśmy analizę jakościową treści, która miała na celu poznanie sposobów komunikowania emocji przez internautów w odniesieniu do Święta Niepodległości.

Adam Pawłowski (Uniwersytet Wrocławski), *Historia i znaczenie badań ilościowych języka dla humanistyki cyfrowej*

Humanistyka cyfrowa jest dziś żywym, interdyscyplinarnym nurtem badań, otwierającym przed nauką nowe perspektywy, związane z możliwościami przetwarzania dużych ilości danych, a także łączenia zasobów językowych (korpusowych) i graficznych. Ale to, co w humanistyce wydaje się nowe, często czerpie inspiracje z przeszłości. Powyższą tezę dobrze ilustrują badania ilościowe języka, rozwijające się długo przed nadejściem ery cyfrowej. W referacie przedstawiona zostanie krótka historia badań ilościowych języka i ich wpływ na dzisiejszą humanistykę cyfrową, a na tym tle – współczesny obraz lingwistyki kwantytatywnej (przede wszystkim, chociaż nie wyłącznie, na przykładzie polszczyzny).

Bogusław Skowronek (Uniwersytet Pedagogiczny w Krakowie), *Nowe media a funkcjonowanie dyskursu społecznego*

Przedmiotem wystąpienia będzie wpływ nowych mediów (głównie mediów społecznościowych) zarówno na strukturę języka (zmiany systemowe, problemy reprezentacji i referencji), jak i na dyskurs społeczny zmiany w definicji medialnych obrazów świata, kreowanie subiektywnej debaty opartej na odrębnych ideologicznie wspólnotach dyskursu wraz z technologicznymi sposobami tworzenia postkomunikacji – media lokacyjne, fakenewsy, zjawisko „postprawdy” itp.

Marek Troszyński (Collegium Civitas), *Czy komputer rozpozna hejtera? Automatyczne przetwarzanie języka dla tekstów z mediów społecznościowych*

W wystąpieniu prezentuję próby wykorzystania automatycznego przetwarzania języka naturalnego (NLP) do oznaczania i klasyfikowania tekstów zebranych z mediów społecznościowych. W przypadku tradycyjnych mediów analiza treści polega na interpretacji znaczeń przez przeszkolony zespół koderów/ anotatorów. Dla setek tysięcy wypowiedzi publikowanych w internecie taki schemat działania wydaje się niemożliwy do zrealizowania. Konieczna staje się automatyzacja złożonego procesu interpretacji tekstu. Czy i na ile można zautomatyzować proces przypisywania sensów do poszczególnych fragmentów tekstu? Czy można skutecznie wykorzystać algorytmy do oznaczania takich konstruktów jak „mowa nienawiści”?

Jacek Warchała (Uniwersytet Śląski), *Dialektyka piśmienności i oralności. Język wobec nowych form transmisji*

Zamiarem referatu jest poruszenie problemu relacji między piśmiennością a oralnością na przykładzie tekstów funkcjonujących w dyskursie publicznym, w Internecie i w mediach – chodzi o tekst pisany, w którym dominują struktury charakterystyczne dla tekstu oralnego. Zostanie w nim podjęta próba zdefiniowania tekstu oralnego, a także ujawnienia kategorii gramatycznych typowych dla mówioności. Będzie również poddany oglądowi wpływ struktury tekstu oralnego na jego sytuacyjność oraz wpływ kodów niewerbalnych na strukturę tekstu oralnego.

Anna Wileczek (Uniwersytet Jana Kochanowskiego), *Kod młodych czy kod młodości? Społeczno-kulturowe konteksty „mediatyacji” młodomowy*

Wystąpienie oparte jest na założeniu, że (cyfrowe) zwyczaje komunikacyjne młodego pokolenia nie tylko odzwierciedlają zmiany społeczno-kulturowe zachodzące we współczesnym świecie, ale także je generują. Warto więc zarysować kierunki i przykłady takiego oddziaływania: od gier z normą językową – po „patojęzyk”, od biolektu – po *kidultmowę*, od wspólnoty dyskursu – po wspólnotę sensu.

Piotr Zbróg (Uniwersytet Jana Kochanowskiego), *Reprezentacja społeczna kultury języka w przestrzeni cyfrowej*

W wystąpieniu zostaną omówione sądy użytkowników internetu na temat poprawności językowej oraz – ogólnie – kultury języka w różnych jej aspektach. Na tej podstawie zostanie zaprezentowana uogólniająca reprezentacja społeczna kultury języka Polaków, czyli swoisty jej obraz społeczny.

Rafał Zimny (Uniwersytet Kazimierza Wielkiego), *Cyfryzacja komunikacji a egalitaryzacja wzorców językowych*

W referacie będą analizowane fenomeny komunikacyjnojęzykowe prymarnie przeznaczone do rozpowszechniania za pomocą nowych mediów (SMS-y, profile i wpisy na portalach społecznościowych, notki blogowe, wiadomości dziennikarskie i inne). Tezą referatu jest to, że cyfryzacja komunikacji międzyludzkiej wpływa współcześnie jako czynnik zasadniczy

na przemiany istniejących wzorców językowych oraz kształtuje nowe. W części analitycznej rozpatrzone zostaną m.in. następujące zjawiska szczegółowe: 1. spłaszczanie się wzorców gatunkowych wysłownia publicznego, 2. wyrównywanie się normy stylowej w kierunku dominacji standardu użytkowego, 3. zanikanie klasowej dystynkcji normy wysokiej, 4. przemiany modelu poradnictwa językowego – dywersyfikacja poziomów kompetencji doradczych (ekspercko-dydaktyczny, ekspercko-profesjonalny, ironiczno-klasowy, partycypacyjno-pluralistyczny). W podsumowaniu referatu zawarte zostaną także wnioski i dezyderaty dotyczące dalszych kierunków działań akademickiego środowiska językoznawczego w zakresie przyszłych standardów kodyfikacji uwzględniających zmiany kulturowo-społeczne wynikające z upowszechnienia się nowych mediów.

ZAJĘCIA WARSZTATOWE

Maciej Piasecki, *CLARIN-PL w warsztacie badacza języka polskiego*

Podczas zajęć zostanie zaprezentowana ogólnoeuropejska infrastruktura naukowa CLARIN ERIC oraz jej integralna część – CLARIN-PL (<http://clarin-pl.eu>). Celem Centrum Technologii Językowych CLARIN-PL jest uzupełnienie luk w systemie tzw. podstawowych narzędzi i zasobów dla języka polskiego. Uczestnicy zostaną zapoznani zarówno z formalnymi, jak i technicznymi możliwościami wykorzystania narzędzi oraz danych udostępnionych w ramach infrastruktury CLARIN-PL.

Marcin Oleksy, *DSpace i CLARIN Cloud – miejsce przechowywania, publikowania i przetwarzania źródeł językowych*

Zajęcia polegają na przeszkoleniu uczestników z korzystania z systemów *DSpace* i *CLARIN Cloud* służących do deponowania i przetwarzania zdeponowanych zasobów językowych. Przedstawione zostaną wzajemne powiązania systemów oraz ich miejsce w międzynarodowej infrastrukturze naukowej CLARIN ERIC.

Zajęcia przeznaczone są dla osób, które planują prowadzenie prac badawczych na korpusie tekstów oraz stworzenie kompleksowego opisu metadanymi grupy dokumentów w celu jego udostępnienia.

Jan Wieczorek, *Cyfrowe źródła leksykograficzne: „Platforma leksykalna” + „Słowosieć”*

Zajęcia mają na celu zapoznanie uczestników z zasobami leksykalnymi języka polskiego i sposobami ich przeszukiwania. *Platforma leksykalna* (<http://lexp.clarin-pl.eu/>) to miejsce, które umożliwia dostęp do danych leksykalnych z różnych źródeł (np. słownik gramatyczny SGJP, Open Multilingual Wordnet, HASK, Słownik XVI w., Słownik XVII w. czy Słownik wileński).

Słowosieć (<http://plwordnet.pwr.wroc.pl/wordnet/>) jest wielkim słownikiem relacyjnym, w którym znaczenia słów (jednostki leksykalne) są opisane za pomocą relacji na poziomie znaczeń i synsetów (zbiorów synonimów).

Zajęcia przeznaczone są dla osób, które planują prowadzić badania w zakresie leksyki synchronicznej języka polskiego na różnych poziomach (semantyki, morfologii, słowotwórstwa, frazeologii), zarówno jeśli chodzi o język ogólny, jak i zawężony do zdefiniowanych obszarów (np. słownictwo specjalistyczne, środowiskowe, ograniczone dyskursywnie).

Maciej Piasecki, *Wielowymiarowa analiza statystyczna tekstów: LEM*

Zajęcia będą polegały na demonstracji działania narzędzia służącego do wydobywania informacji z tekstów. Uczestnicy warsztatów zostaną zapoznani z możliwościami zastosowania systemu LEM (<http://ws.clarin-pl.eu/lem.shtml>), który umożliwi m.in.: lematyzację, wyznaczanie części mowy i nazw własnych, charakterystykę czasownikowej, generowanie statystyki słów i części mowy, ujednoznacznianie znaczeń leksykalnych oraz analizę wydźwięku emocjonalnego.

Marcin Oleksy, *Narzędzia elementarne do przetwarzania tekstów – przewodnik (segmentacja, analiza morfosyntaktyczna, rozpoznawanie znaczeń słów, rozpoznawanie nazw własnych, wyrażen temporalnych i przestrzennych...)*

Przedmiotem zajęć będą serwisy webowe umożliwiające automatyczne przetwarzanie tekstów. Uczestnicy zostaną zapoznani z podstawami teoretycznymi, punktami dostępu oraz sposobem wykorzystania narzędzi NLP, które poddają teksty automatycznej analizie w zakresie segmentacji, analizy morfosyntaktycznej, rozpoznawania znaczeń słów, nazw własnych, wyrażen temporalnych i przestrzennych.

Jan Wieczorek, *Systemy analizy semantycznej tekstów (wydźwięk emocjonalny, słowa kluczowe...)*

Podczas zajęć zostaną zademonstrowane najważniejsze funkcje systemów przeznaczonych do analizy semantycznej tekstów. Przy ich użyciu możliwe jest określenie wartości wydźwięku emocjonalnego, słów kluczowych, występowania fraz odnoszących się do relacji przestrzennych i czasowych oraz występowania wtrąceń obcojęzycznych. Systemy tego typu mogą znacznie przyspieszyć badania naukowe w zakresie dyscyplin filologicznych, są również przyjazne w obsłudze dla osób bez szerszego zaplecza informatycznego.

Jan Wieczorek, *Badania humanistyczne z wykorzystaniem narzędzi do automatycznego przetwarzania języka polskiego – przewodnik*

Podczas zajęć zostaną zaprezentowane wnioski płynące z wieloletniej współpracy konsorcjum CLARIN-PL z różnymi zespołami badawczymi realizującymi projekty naukowe z zakresu nauk humanistycznych i społecznych. Wnioski te dotyczą sposobów projektowania badań wykorzystujących automatyczne przetwarzanie języka: sugerowanych etapów przygotowań, doboru właściwych narzędzi, metod unikania najczęstszych błędów. Zajęcia przeznaczone są dla osób o niewielkim lub żadnym doświadczeniu w podobnej działalności, chcących podjąć badania z zakresu humanistyki cyfrowej.

Jan Wieczorek, *Przykłady zastosowań technologii cyfrowych we współczesnych badaniach nad językiem polskim*

Podczas zajęć zostaną zaprezentowane i omówione zakończone lub trwające badania naukowe prowadzone przy zastosowaniu technologii automatycznego przetwarzania języka. Zajęcia mogą stanowić inspirację do zaprojektowania własnych badań oraz przegląd typowych problemów, z którymi spotyka się badacz działający na polu humanistyki cyfrowej.

Marcin Oleksy, *Tworzenie własnych cyfrowych korpusów językowych*

Przedmiotem zajęć będzie tworzenie cyfrowych korpusów językowych na podstawie danych zgromadzonych przez badacza. Uczestnicy warsztatów zostaną zapoznani z systemem Inforex (<https://inforex.clarin-pl.eu/>) oraz możliwościami, jakie on daje w zakresie: konstruowania korpusu i zarządzania pracą oraz generowania i wykorzystania prostych statystyk dotyczących badanych tekstów (np. listy frekwencyjne wyrazów i lematów, statystyki dotyczące kolokacji i konkordancji). Omówione zostaną również powiązania pomiędzy Inforexem a systemami do deponowania danych – *DSpace* i *CLARIN Cloud*.

Marcin Oleksy, *Ręczna i automatyczna anotacja korpusów*

Przedmiotem zajęć będzie system do anotacji tekstów – *Inforex* oraz połączone z nim systemy *CLARIN Cloud* i *DSpace*. Zaprezentowane zostaną funkcjonalności tych narzędzi dotyczące zarówno ręcznego, jak i automatycznego znakowania (np. w zakresie znaczników morfosyntaktycznych, nazw własnych i wyrażen temporalnych). Zajęcia przeznaczone są dla osób, które planują w swojej pracy badawczej przeprowadzenie anotacji (znakowania, tagowania) tekstów za pomocą zdefiniowanej przez siebie warstwy anotacji oraz pracę z danymi zidentyfikowanymi w procesie znakowania.

Tomasz Walkowiak, *System analizy stylu i podobieństwa tekstów: „WebSty”*

Demonstracja działania i wykorzystania narzędzia służącego do badania parametrów stylometrycznych. Uczestnicy warsztatów zostaną zapoznani z podstawami teoretycznymi i możliwościami zastosowania systemu WebSty (<https://ws.clarin-pl.eu/websty.shtml>), który umożliwia przeprowadzenie analizy tematycznej badanych tekstów, grupowanie tematyczne i stylistyczne tekstów, analizę autorstwa i analizę stylu gramatycznego.