

Tekstynų analizės programa UNITEX ir jos taikymo galimybės

Skaistė Volungevičienė, Alina Isilionis

Vokiečių filologijos katedra

Vilniaus universitetas

Universiteto g. 5

LT-01513 Vilnius, Lietuva

Tel.: +370 5 2687230

El. paštas: skaiste.zabarauskaite@flf.vu.lt, alina.isilionis@flf.stud.vu.lt

Anotacija

Paskutiniaisiais dešimtmečiais kompiuteris tapo daugelio lingvistinių tyrimų instrumentu. Šiuolaikinių tyrimų kalbininkai neįsivaizduoja be kalbinės programinės įrangos, kurios pasaulyje sukurta daug ir įvairios, tačiau pasinaudoti kitų sukurtais produktais dėl įvairių priežasčių tyrėjams pavyksta ne visada. Šiame straipsnyje pristatoma tekstynų analizės programa UNITEX, kurios idėja kilo ir buvo realizuota Prancūzijos LADL laboratorijos (*Laboratoire d'Automatique Documentaire et Linguistique*) profesoriaus Maurice Grosso. Iš pradžių programa UNITEX buvo pritaikyta analizuoti prancūzų kalbos tekstynus, bet vėliau jos galimybės išplėstos, o po paskutinio programos naujinimo šiuo įrankiu galima apdoroti net 17 kalbų tekstus. Nors šių kalbų sąrašė lietuvių kalbos kol kas nėra, ši programa, straipsnio autorių manymu, galėtų sudominti Lietuvos tyrėjus ir paskatinti ieškoti galimybių taikyti ją lietuvių kalbos analizei. Pirmojoje šio straipsnio dalyje trumpai pristatomos programos UNITEX galimybės, o antrojoje išsamiau aprašoma viena šios programos praktinio taikymo sričių – autorių sukurti vokiečių kalbos verbonominalinių konstrukcijų paieškos algoritmai.

Raktažodžiai: tekstynų analizės įrankiai, UNITEX, lokalsiosios gramatikos, paieškos algoritmai, verbonominalinės konstrukcijos

1 Įžanginės pastabos

Nors moderniais, kitose šalyse ilgametę taikymo tradiciją turinčiais kompiuterinės lingvistikos tyrimų metodais Lietuvoje susidomėta gana neseniai, tačiau jau prieš dešimtmetį publikuotuose straipsniuose akcentuojama, kad programiniai įrankiai – neatskiriama lingvistinių tyrimų dalis. „Kompiuteriai ir kompiuterinės technologijos

atvėrė naujas galimybes, parodė dažnumus, reguliarius kalbos vartosenos modelius.“ (Marcinkevičienė 2000, 37) Net ir tradicinių darbo metodų šalininkas sutiktų, kad kompiuteris bei įvairi programinė įranga ne tik palengvina mokslininko darbą, bet ir suteikia galimybių į analizuojamą kalbinį reiškinį žvelgti daug plačiau ir taupyti laiką. Šio straipsnio tikslas – pristatyti Lietuvoje dar nenaudojamą tekstynų analizės programą UNITEX, jos technines galimybes bei pateikti praktinių šios programos taikymo pavyzdžių. Dėl ribotos straipsnio apimties išsamiai aprašyti visų programos funkcijų, žinoma, neįmanoma, tačiau tai ir nėra autorių tikslas. Autorės siekia sudominti tyrėjus, paskatinti juos tobulinti ir lietuvių kalbai pritaikyti jau sukurtus analizės įrankius, o ne pateikti išsamų UNITEX vartotojo vadovą. Toks vadovas būtų aktualus ir būtinas, jei UNITEX analizuojamų kalbų sąrašė ateityje atsirastų ir lietuvių kalba, o šalies kalbininkai matytų galimybę šį programinį įrankį taikyti savo praktiniuose tyrimuose.

Šiuo metu Lietuvoje pastebimos tendencijos kurti naujus, o ne bandyti pritaikyti jau egzistuojančius produktus: savarankiškai sukurtas ir toliau tobulinamas lietuvių kalbos klaidų tikrintuvai „Juodos avys“, kuris remiasi *Dabartinės lietuvių kalbos žodynu* (1972), *Tarptautinių žodžių žodynu* (1982) ir *Lietuvių kalbos gramatika. Fonetika ir morfologija* (1965, 1971), skiemenuoklis „Skie-muo“, morfologinei analizei reikalingas įrankis „Lemuoklis“, „sukurtas veiksmažodžių grupių analizatorius ateityje bus pritaikytas priklausomybių gramatika paremtam sintaksiniam lietuvių kalbos analizatoriui kurti“ (Grigonytė, Rimkutė 2005, 315) ir t. t. Kitų kalbų tekstynų sudarytojai ir mokslininkai elgiasi kitaip: jie pirmiausia ieško poreikius atitinkančių programinių įrankių, kai tokių neranda, bando pritaikyti jau egzistuojančius produktus ir tik nepasisekus – kuria naujus. Šiuo atveju argumentas, kad analizuoti didžiąsias Europos kalbas dėl jų struktūros panašumo tinka arba analizei atlikti lengvai pritaikomos visos sukurtos priemonės, kurias pritaikyti lietuvių kalbai yra labai sunku, skamba įtikinamai, tačiau geriau su jomis susipažinus ir pastudijavus jų veikimo principus tenka pripažinti, kad tai yra įmanoma. Suprantama, norint taikyti kompiuterines kalbos analizės priemones vis naujų kalbų analizei, reikia daugybės kompleksinių, kruopščiai atliktų kalbos tyrimų, kurių Lietuvoje vis dar trūksta. Taip pat dažnai nepakanka ir žinių apie tai, kokių produktų jau yra sukurta.

2 Techninės UNITEX galimybės

Vienas dažniausių kalbos analizei skirtų kompiuterinių priemonių grupavimo kriterijų – jų atliekamos funkcijos ir paskirtis. Pagal šį kriterijų skiriami konverteriai, dažninių sąrašų generatoriai, konkordavimo programos, sintaksinės analizės programos, klaidų tikrintuvai, skiemenuokliai, lemuokliai, morfologiniai analizatoriai, anotatoriai,

paralelinimo programos, leksinės duomenų bazės ir mašininio vertimo sistemos (Utka 2000, 275–277). Reikia pastebėti, kad programa UNITEX nėra siauros paskirties įrankis, greičiau ji yra įrankių rinkinys, atliekantis keletą funkcijų. Programa leidžia konvertuoti grynojo teksto rinkmenas į tolesnei analizei reikalingą *snt dokumentų formatą, gauti dažninius analizuojamo tekstyno žodžių sąrašus bei paieškos sąlygas atitinkančius konkordansus, taip pat kurti paieškos algoritmus. Naudojantis UNITEX galima atlikti morfologinę, sintaksinę bei leksinę tekstyno analizę. Ši programa padeda kurti naujus kalbos analizei skirtus įrankius, internetinės paieškos sistemas, mašininio vertimo programas ar rašybos ir gramatikos tikrintuvus.

Šiuo metu didžioji dalis tekstynų tyrimų atliekami naudojant universalų visoms operacinėms sistemoms pritaikytą programinių įrankių paketą *WordSmith Tools*, tačiau pagrindinis jo minusas – gana didelė kaina, todėl dažnas vartotojas ieško panašias funkcijas atliekančio nemokamo įrankio. Iš nemokamų produktų reikia paminėti vis labiau populiarėjančius *AntConc* ir *TextSTAT*. Šie įrankiai taip pat veikia visose operacinėse sistemose, gali atpažinti įvairių kalbų, tarp jų ir lietuvių kalbos, tekstus, tačiau tai siauros paskirties įrankiai, leidžiantys tik generuoti dažninius sąrašus ar gauti paieškos kriterijus atitinkančius konkordansus. Norėdami pasinaudoti paieškos algoritmų kūrimo funkcija turinčiais programiniais įrankiais dažniausiai turime mokėti bent vieną iš programavimo kalbų (pvz.: *Perl*, *Python*, *C++* ar kt.). Vis labiau populiarėjantis, nemokamas, lingvistinėms užduotims spręsti tinkamas produktas – programa *R*. Tiesa, kalbininkams juo naudotis nėra paprasta, nes būtina išmanyti ne tik programavimo, bet ir statistikos subtilybes. Naudojantis autorių pristatomu įrankiu UNITEX, programavimo kalbos mokėjimas nėra būtinybė, o algoritmų kūrimo kalba labai paprasta ir greitai išmokstama.

Programa UNITEX gimė viename Prancūzijos mokslinių tyrimų centre (*Labaratoire d'Automatique documentaire et Linguistique*). 2002 m. tyrėjams buvo pristatytas naujas, visoms operacinėms sistemoms (*Windows*, *Linux*, *MacOS*) pritaikytas, paplitusioje *Java* programinėje terpėje sukurtas programinis įrankis UNITEX, laisvai prieinamas¹ visiems vartotojams. Šis įrankis tapo atsvara iki tol tyrimams naudotai komercinei programai INTEX. Iš pradžių programa UNITEX buvo galima analizuoti tik prancūzų kalbos tekstus, tačiau dabar programa yra išstbulinta ir naujausia jos versija UNITEX 2.1 „supranta“ net 17 kalbų: anglų, senąją ir naująją graikų, ispanų, italų, korėjiečių, lenkų, norvegų, prancūzų, portugalų (brazilų ir portugalų variantus), rusų, serbų (kirilica ir lotyniški rašmenys), suomių, vokiečių ir tajų). Šią kalbų įvairovę atspindi ir programos

¹ Programą galima parsisiųsti iš oficialaus UNITEX puslapio <http://www-igm.univ-mlv.fr/~unitex/index.php?page=3>.

logotipas, kuriame pavaizduota 12 skirtingų specialių analizuojamųjų kalbų ženklų (žr. 1 pav.).

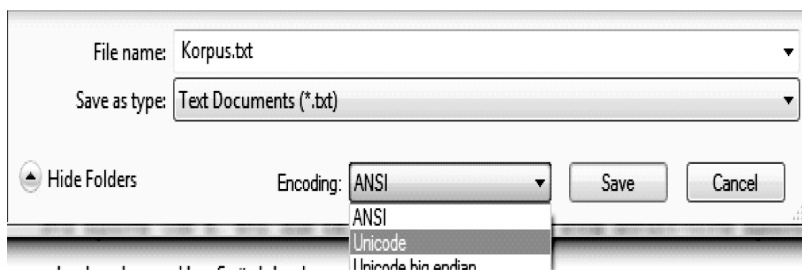


1 pav. Programos UNITEX logotipas

Dėl ribotos straipsnio apimties toliau pristatomos tik pagrindinės programos funkcijos. Daugiau informacijos apie visas UNITEX galimybes galima rasti Paumiero (2008) parengtame vartotojo vadove.

2.1 Tekstyno kūrimas programa UNITEX

Norėdami pradėti dirbti programa visada turime pasirinkti kalbą. Jei prireikia, kalbą bet kada galime pakeisti naudodami meniu punktą *Text* → *Change Language...*. Kad UNITEX galėtume kurti tekstynus, tekstai turi būti išsaugoti grynojo teksto formatu, t. y. tyrėjui reikalinga viena ar kelios rinkmenos su *.txt plėtiniu. Išsaugant rinkmenas svarbu pasirinkti tinkamą koduotę. Standartiškai tekstą siūloma išsaugoti koduotėje ANSI, tačiau norint dirbti su programa UNITEX reikia UNICODE standarto, leidžiančio apdoroti specialiųjų ženklų turinčius tekstus, pvz., sudaryti Azijos kalbų tekstyną. Nepasirinkę standarto UNICODE, kaip parodyta 2 pav., su tyrimui parengtu tekstynu toliau dirbti negalėsime.

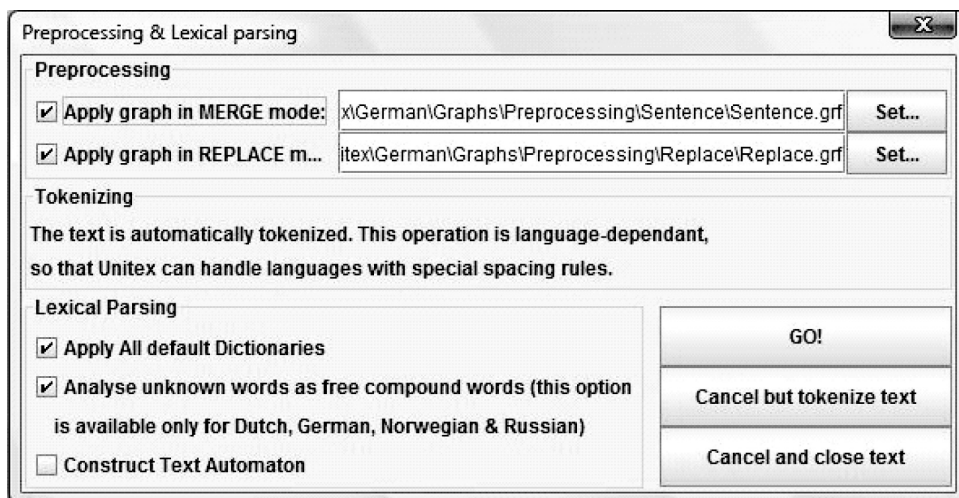


2 pav. Grynojo teksto dokumento koduotės pasirinkimas

Norėdami dirbti su tekstynu išsaugotą grynojo teksto rinkmeną turime atverti UNITEX, t. y. pasirinkti meniu punktą *Text* → *Open...*, jei šią rinkmeną atidarome pirmą kartą,

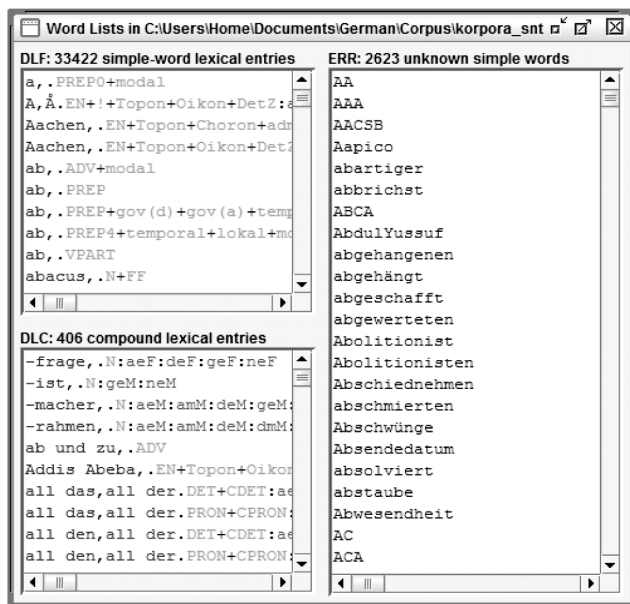
arba *Text* → *Open Tagged Text...*, jei naudojames programos jau apdorotu dokumentu. Pasirinkus atverti nauja tekstą, pasirodo užklauso langas (žr. 3 pav.), kuriame programai nurodome, ką reikia atlikti pradiniu teksto apdorojimo etapu.

Vartotojo pageidavimu vyksta pirminis teksto apdorojimas: tekstas paruošiamas tolesniam darbui, t. y. išvalomas nuo tarpų, eilučių lūžių, kurie atsiranda, kai tyrimui reikalingi tekstai renkami iš interneto. Taip pat programa UNITEX pagal programos kūrėjų kiekvienai kalbai parengtas taisykles nustato sakinių ribas – kiekvieno sakinio pabaigoje atsiranda ženklas {S}. Apdorojant tekstą pirmą kartą gali būti analizuojami ir kontrakcijos atvejai, pvz., anglų kalbos *I'm* virsta *I am*, vokiečių *geht's* keičiama į *geht es* ir t. t. Galimi kiekvienos kalbos kontrakcijos atvejai yra numatyti programos kūrėjų sukurtuose, programoje įdiegtuose algoritmuose. Vėliau, atlikus šiuos formaliuosius keitimus, tekstas skaidomas į leksemas ir programa, naudodamasi įdiegtais žodynais, atlieka automatinę leksinę analizę. Programa UNITEX anotuoja tekstyną pagal tai, ką ir kaip kategorizavo tam tikros kalbos žodyno sudarytojai, taigi, norint sėkmingai dirbti šia programa pirmiausiai reikia parengti tokį žodyną, kurį vartotojai vėliau gali pildyti pagal savo poreikius. Deja, reikia pripažinti, kad esminis šio programinio įrankio trūkumas – to negali padaryti paprastas vartotojas. Jei Lietuvos mokslininkai susidomėtų UNITEX, vien tik rengėjų sutikimo naudoti, pvz., *Dabartinės lietuvių kalbos žodyną* neužtektų, tačiau kiek straipsnio autorėms teko bendrauti su bendruose projektuose dirbančiais Miuncheno Liudvigo Maksimiliano universiteto (Vokietija) Informacijos ir kalbų apdorojimo centro (CIS) darbuotojais, susidomėjimas mažomis retomis kalbomis yra didelis ir Prancūzijos LADL laboratorija tikrai būtų pasirėngusi UNITEX apdorojamų kalbų sąrašą papildyti nauja kalba.



3 pav. Pradinio teksto apdorojimo etapo užklauso langas

Gauti automatinės leksinės analizės rezultatai pateikiami atskirame 3 dalių lange: vienoje dalyje rodomos anotuotos sudaryto teksto leksemos, antroje dalyje pateikiamos keliažodės samplaikos, o trečioje – programoje įdiegtame žodyne nefiksuotos leksemos (žr. 4 pav.).

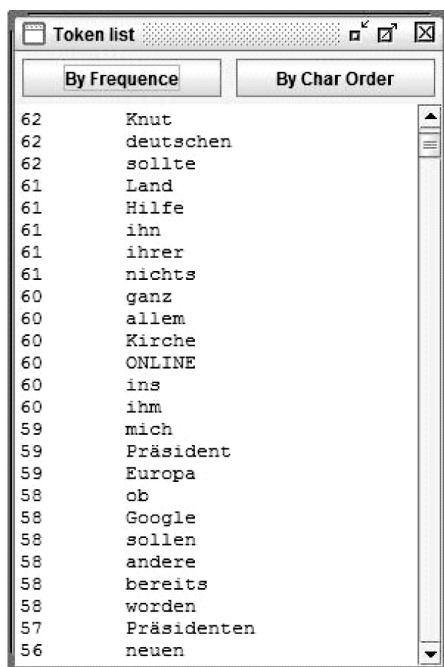


4 pav. Leksinės analizės rezultatų langas

Kaip matyti iš 4 pav., leksinės analizės rezultatų lange pateikiama tekстыne rasta žodžio forma ir jo pagrindinė forma, jeigu ji skiriasi nuo rastosios, taip pat nurodoma kalbos dalis, fiksuojama gramatinė bei semantinė informacija. Jei reikalinga, prie leksemų gali būti pateikta komentarų².

Pirminiu teksto apdorojimo etapu gaunamas ir dažninis analizuojamo teksto žodžių sąrašas (žr. 5 pav.), kurį programa leidžia rikiuoti pagal žodžių dažnumą arba abėcėlės tvarka. Po šio etapo programa grynojo teksto dokumentą automatiškai išsaugo *snt formatu, kurį reikia naudoti dirbant toliau, nes tai vienintelis programos atpažįstamas formatas.

² Informacija apie leksemas programoje koduojama spalvomis: mėlyna spalva žymima tekстыne rasta žodžio forma, raudona – pagrindinė jo forma, žalia spalva koduojama gramatinė, o ruda – semantinė informacija. Kadangi šis leidinys nespaltotas, 4 pav. to nematyti.

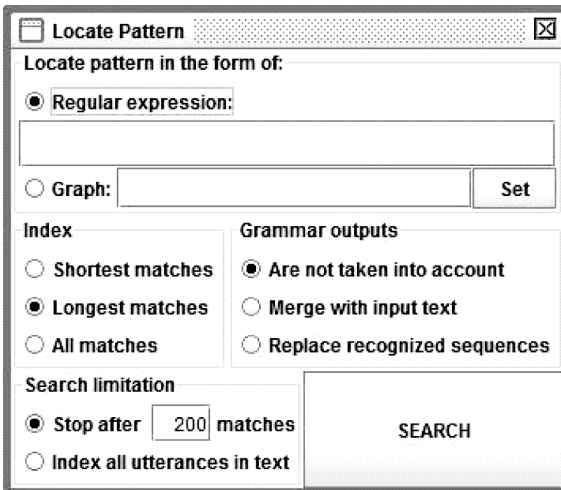


5 pav. Dažninis tekstyne pasitaikančių žodžių sąrašas

2.2 UNITEX paieška

Programoje UNITEX visa informacija yra koduojama, t. y. paieškai ir algoritmams kurti naudojamos taip vadinamosios „kaukės“ – tam tikros duomenų koduotės, taikomos kitai tokios pat struktūros reikšmei identifikuoti, kai siekiama iš tekstyne išskirti tam tikrą duomenų dalį. Šioje programoje naudojamos 3 rūšių kaukės: leksinės, morfologinės ir meta-kaukės; taip pat UNITEX leidžia paieškai naudoti įvairias kaukių kombinacijas. Norėdami susikurtame tekstyne vykdyti paiešką, galime rinktis programos meniu punktą *Text* → *Locate Pattern...* arba kompiuterio klavišų kombinaciją *Ctrl + L*. Abiem atvejais ekrane pasirodo 6 pav. pavaizduotas paieškos užklauso langas. Dominančios struktūros tekstyne ieškoma naudojantis paieškos kaukėmis (žymimas punktas *Regular expression*) arba pačių ar kitų sukurtais paieškos algoritmais (renkamės punktą *Graph*). Kai tekstinys yra labai didelis, šiame paieškos užklauso lange paieškos rezultatų skaičių galime apriboti, pvz., nurodyti, kad programa pateiktų tik pirmus 200 paieškos užklausa atitinkančių rezultatų.

Leksinės paieškos kaukės naudojamos tada, kai tyrėjui reikalingi tekstyne leksinės analizės duomenys, t. y. kai mus domina paradigminės konkrečios leksemos formos, tam



6 pav. Paieškos užklauskos langas

tikra kalbos dalis ar net leksinės samplaikos. Meta-kaukės pritaikytos atlikti formaliųjų dalykų paiešką, pvz., ieškoti tekстыne pasitaikančių skaičių kombinacijų, žodžių, kurie prasideda mažąja ar didžiąja raide (tai pvz. gali būti svarbu tiriant vokiečių kalbą, kurios daiktavardžiai rašomi didžiąja raide) ir pan. Morfologinių kaukių pririekia rūšiuojant duomenis pagal žodžių darybos segmentus. Pirmojoje lentelėje pateikti keli paieškos kaukių pavyzdžiai (išsamiau žr. Paumier 2008, 71–89).

Leksinės kaukės	<leksema>	visos paradigminės formos
	<leksema.V/N/A/ADV/PRON>	visos paradigminės tam tikros kalbos dalies formos
	<CDIC>	leksinės samplaikos
Meta-kaukės	<MIN>	žodžiai mažosiomis raidėmis
	<MAJ>	žodžiai didžiosiomis raidėmis
	<PRE>	žodžiai, prasidedantys didžiąja raide
	<NB>	skaičių kombinacijos
Morfologinės kaukės	<<ss>>	žodžiai, kuriuose pasitaiko 'ss'
	<<^a>>	žodžiai, prasidedantys 'a'
	<<t\$>>	žodžiai, kurie baigiasi 't'
Kaukių kombinacijos	<CDIC><<->>	sudurtiniai žodžiai, kuriuose yra '-'
	<.N+HUM>	daiktavardžiai, turintys semą + <i>Hum</i>

1 lentelė. Paieškos kaukių pavyzdžiai

Baigusi paiešką programa pateikia trumpą informaciją apie jos rezultatus, t. y. praneša, kiek rasta paieškos kriterijus atitinkančių įrašų, ir siūlo sudaryti jų konkordansą, kuris generuojamas paspaudus mygtuką *Build concordance*. Toliau 7 pav. pateiktas konkordanso, gauto paieškai panaudojus kaukių kombinaciją <CDIC><<->>, pavyzdys.

die ersten Erfahrungen aus den USA mit E-Mail-Fundraising vor.{S} Die Ergebnisse zeigen, dass Beratung: Opfer und Strafrecht per Tel./E-Mail; 8-24h{S} www.deutsche-anwaltshotline.de{S} ----er beraten{S} Die Verkehrsminister der EU-Länder wollen über das durch die Aschewolke aus Isla S) Trotz teils heftiger Kritik einiger EU-Länder spielt die deutsche Regierung auf Zeit: Bunde im März 1999 lebte eine bedeutende{S} kosovo-albanische Gemeinschaft in der Schweiz.{S} Desha weil der Bund und die Länder Berlin und Mecklenburg-Vorpommern eine Ausfallbürgschaft über 146 nne-Katrin Mellmann, ARD-Hörfunkstudio Mexiko-Stadt{S} Tagelang haben die meisten Einwohner vo rauischen{S} Anzeige{S} Thema: Wahl in Nordrhein-Westfalen{S} Spanien{S} Solidarität für Balta ausstieg Kurz vor der Landtagswahl in Nordrhein-Westfalen nutzten sie die Gelegenheit zum dem vor dem Atommüll-Zwischenlager Ahaus in Nordrhein-Westfalen.{S} Anlass für den Protest war der Element.{S} Ihre alte Partei sammelt in Nordrhein-Westfalen Unterschriften gegen diese Pauschal habe damals die Landesvertretung von Nordrhein-Westfalen in Brüssel aufgebaut.{S} Das gescha des Gesicht gaben.{S} Wo ist jetzt der nordrhein-westfälische Integrationsminister Armin Lasch mit Tränengas.{S} Die Bevölkerung von Port-au-Prince, der Hauptstadt Haitis, wird ungeduldig. gelang haben die meisten Einwohner von Port-au-Prince vergeblich auf Hilfe gewartet.{S} Aufgeb einem Land wie Österreich, in dem die römisch-katholische Tradition über Jahrhunderte prägend das jener Mann, der an der Spitze der römisch-katholischen Kirche steht und einst als Chef de re laufenWährend der Regierungszeit von Rot-Grün hat der Bundestag den Atomausstieg beschlossen ein ans Netz.{S} Mit dem Beschluss der rot-grünen Bundesregierung im Jahr 2000, bis 2021 aus d ung hat unter anderem die Regierung von Sachsen-Anhalt angekündigt.{S} Weiterführende Links{S}

7 pav. Konkordanso pavyzdys

2.3 Algoritmų kūrimas

Kita labai svarbi pristatomos programos funkcija – galimybė kurti algoritmus. Gerai parašytas paieškos algoritmas tiksliai atspindi realią kalbos vartojimo situaciją ir sutaupo nemažai laiko. Kuo geresnis algoritmas, tuo daugiau jis apima įvairių frazių ar sakinio konstrukcijų, atpažįsta struktūrų ir tuo mažiau pasitaiko atpažinimo klaidų, pvz., naudojant prastą algoritmą, atpažįstama tik dalis reikalingos struktūros arba ji visai neaptinkama, nors mokslininkas ir įsitikinęs, kad tiriamame tekстыne struktūra egzistuoja. Kad vartotojas išsivaizduotų, kaip atrodo ir kokios apimties algoritmų gali būti, programos UNITEX vokiečių kalbos dalyje jau įdiegti du Sebastiano Nagelio (2005) sukurti algoritmai: vienas jų skirtas daiktavardinėms frazėms, o kitas – prielinksninėms konstrukcijoms atpažinti.

Norėdami naudotis programos UNITEX algoritmų kūrimo funkcija, renkamės meniu punktą *FSGraph* → *New...* ir programa atveria specialų algoritmų kūrimo langą. Kiekvienas algoritmas turi rodykle žymimą pradžios tašką ir kvadratėliu apskritime žymimą pabaigos tašką (žr. 8 pav.). Algoritmas kuriamas ar interpretuojamas iš kairės į dešinę. Visi algoritmą sudarantys segmentai išdėstomi tarp šių dviejų taškų ir sujungiami.

Vieno algoritmo segmentų skaičius neribojamas: kuo jų daugiau, tuo atpažįstama struktūra yra kompleksiškesnė. Algoritmams kurti naudojamos tos pačios duomenų kaudių kombinacijos kaip ir paprastoms paieškoms.



8 pav. Algoritmų kūrimo langas

Išbandyti programą UNITEK kuriant verbonominalinių konstrukcijų paieškos algoritmus pasirinkta dėl aiškios ir stabilios šių konstrukcijų struktūros bei dažno vartojimo. Ne mažiau svarbus pasirinkimo argumentas: didelis mokslininkų susidomėjimas šiuo kalbiniu reiškiniu. Verbonominalines konstrukcijas išsamiai tiria įvairių šalių mokslininkai, norėdami sukurti su jomis galinčią susidoroti mašininio vertimo sistemą. Kadangi straipsnio autorės vertimo problemos taip pat domina, o lietuvių tyrėjai verbonominalines konstrukcijas yra menkai tyrę, jos atsirado šio tyrimo dėmesio centre.

3 Verbonominalinės konstrukcijos

Jeigu atliktume apklausą, tikriausiai daugelis tvirtintų, kad iš bendros ar išblukusios reikšmės veiksmažodžių ir veiksmažodinių daiktavardžių sudarytos konstrukcijos lietuvių kalbai nebūdingos, netgi manoma, kad yra taisytinės, turbūt todėl šiam reiškiniui lietuvių kalbotyroje daug dėmesio ir neskiriama. Tą įrodo ir faktas, kad lietuvių kalboje nėra nusistovėjusio termino tokio tipo junginiams pavadinti: kartais lingvistai rašo apie junginius su funkciniais veiksmažodžiais (Alosevičienė 2004) ar analizuoja veiksmažodį atitinkančias konstrukcijas (Kniūkšta 2005). Šiame darbe, remiantis vokiečių tradicija, tokias konstrukcijas siūloma vadinti verbonominalinėmis. Užsienio lingvistų darbuose joms skiriama labai daug dėmesio. Išsamiai nagrinėjami anglų, prancūzų, vokiečių, olandų ir kt. kalbų junginiai, konstatuojama, kad verbonominalinės konstrukcijos jose

yra dažnos, bet dėl vieningo termino taip pat nesutariama: vietoje anksčiau prancūzų tradiciškai vartoto termino *locutions verbales* įsitvirtino *constructions à verbe support*, anglakalbėje literatūroje galima rasti *light verb*, *complex predicate* arba *phrasal verb*, *support verb* ir *delexical verb*, vokiečių lingvistams gerai žinomas *Funktionsverbgefüge* terminas, nors naujausiuose darbuose vis dažniau pasitaiko *Stützverbkonstruktion*, kuris atsirado į vokiečių kalbą išvertus prancūzų *constructions à verbe support*. Į klausimą, ar verbonominalines konstrukcijas galima traktuoti kaip universalų fenomeną, Kōrōsi atsako apžvelgusi skirtingų tipų kalbų tyrimus. Teigiamą atsakymą ji patvirtina agliutinacinių, izoliacinių, fleksinių ir inkorporacinių kalbų tyrimų pavyzdžiais. Pagrindinė sąlyga – kalbose turi egzistuoti daiktavardžio ir veiksmažodžio kategorijos (Kōrōsi 2008, 96–97).

Nors kiekvienas verbonominalinių konstrukcijų fenomeną tyręs mokslininkas bandė įvardinti kuo daugiau semantinių ir formaliųjų kriterijų, leidžiančių nubrėžti ribą tarp išblukusios reikšmės veiksmažodžių ir veiksmažodinių daiktavardžių junginių ir kitos struktūros junginių, tačiau vieno apibrėžimo iki šiol nėra, o dažnas tyrėjas savo teorinius apmąstymus baigia mintimi, kad visi minėti kriterijai „dažniausiai galioja ne visiems šios grupės atstovams“ (Hentschel, Weydt 1990, 79). Užduotis pasunkėja dar ir todėl, kad nagrinėjamas kalbos fenomenas atsiduria paribyje tarp morfologijos ir sintaksės, o jam apibrėžti pasitelkiami tokie pat neaiškūs ir diskutuoti kitių leksinių vienetų apibrėžimai. Kas vienam mokslininkui neginčijamai priklauso verbonominalinių konstrukcijų būriui, kitų manymu, į tokį sąrašą nepatenka (plg. Daniels 1963, 25 ir Engelen 1968, 293 ir 295). Peržvelgus šį fenomeną aprašančius ir nagrinėjančius darbus galima konstatuoti, kad lingvistai sutaria dėl formaliosios junginio struktūros (veiksmažodis + (prielinksnis) + daiktavardis) ir pripažįsta, kad veiksmažodžio semantinė reikšmė tokiuose junginiuose yra išblukusi. Veiksmažodis dažnai atlieka tik gramatinę funkciją, todėl daugelio vadinamas semantiškai tuščiu, nors jis nėra visai tuščias, o perima bendrąsias semantines funkcijas, t. y. išreiškia būseną, jos pakitimą ar pakitimo priežastį (Helbig, Buscha 2001, 70). Nesiginčijama ir dėl teiginio, kad pagrindinis semantinis tokio junginio krūvis tenka daiktavardžiui, kuris yra veiksmažodžio, rečiau būdvardžio vedinys arba *nomen actionis*, o abu junginio elementai sudaro nedalomą semantinį vienetą, sakinyje atliekantį tarinio funkciją. „Kitais atvejais minimos „dažnai“ arba „dažniausiai“ pasitaikančios savybės, labai nesigilinant, kiek jos yra būdingos ar svarbios *Funktionsverbgefüge*.“ (Pottelberge 2001, 2) Dalį tokių junginių dažnai galima pakeisti veiksmažodžiu. Kōrōsi (2008, 29–30) nurodo net 6 jų substitucijos galimybes:

1. junginys keičiamas veiksmažodžiu, į kurį transformuojamas daiktavardis (*zum Ausdruck bringen* → *ausdrücken* ‘išreikšti, pasisakyti’³);

³ Čia ir toliau pateikiami žodyniniai vokiškų konstrukcijų atitikmenys, tačiau konkrečioje kalbinėje situacijoje bendros ar išblukusios reikšmės veiksmažodžių ir veiksmažodinių daiktavardžių junginiai į lietuvių kalbą gali būti/ yra verčiami visiškai kitaip.

2. junginį keičiantis veiksmažodis atitinka ne tik jo daiktavardį, bet visą junginį (*Anspruch erheben* → *beanspruchen* ‘reikalauti’);
3. junginys keičiamas pasyvine konstrukcija (*Erwähnung finden* → *erwähnt werden* ‘būti paminėtam’);
4. junginys keičiamas konstrukcija su veiksmažodžiu *sein* ‘būti’ (*die Fähigkeit besitzen* → *fähig sein* ‘gebėti’);
5. junginys neturi atitiktens (*in Frage kommen* → \emptyset ‘būti susijusiam’);
6. junginį atitinka su jo daiktavardžiu etimologiškai nesusijęs veiksmažodis (*zur Welt bringen* → *gebären* ‘pagimdyti’).

4 Verbonominalių konstrukcijų paieškos algoritmai

Didžioji dalis bandymų siekiant sukurti nuoseklias lingvistines teorijas, galinčias paaiškinti, kaip funkcionuoja kalba, ar ir kaip yra susiję sintaksės, morfologijos ir semantikos lygmenys, buvo nesėkmingi. Net ir pati išsamiausia gramatika negali aprašyti visų įmanomų kalbos vartojimo atvejų. Tą pastebėjo jau Chomskis (1957), bandęs sukurti formaliąsias gramatikas, bet jo sukurtą generatyvinę gramatiką tekstynų lingvistikos specialistai kritikuoja dėl daugybės išimčių. Programą UNITEX kūręs Grossas buvo įsitikinęs, kad gramatikos specialiai konstruoti nereikia, ji egzistuoja konkrečiame tekste, ir prieš kuriant lingvistines teorijas reikia sukaupti pakankamai didelį tekstyną bei ištirti ir aprašyti jame pasitaikančius atvejus. Tiriant tekstynus atliktos sintaksinių ir semantinių struktūrų analizės rezultatas – kalbai aprašyti reikalingų taisyklių rinkinys, kurį Grossas pavadino *lokale Grammatik* terminu (Gross 1997). Rusų lingvistas Mel’čukas šias gramatikas vadino *kalbos žemėlapiais*, aprašančiais kompleksines struktūras, kurių negalima fiksuoti žodynuose (Mel’čuk 2004), t. y. sintaksinio ar semantinio lygmens vienetus. Lietuvių kalboje šiam reiškiniui pavadinti termino kol kas nėra, todėl šiame straipsnyje siūlomas lokaliųjų gramatikų terminas. Šios gramatikos padeda atskleisti žodžių daugiaprasmiškumą, išaiškinantį tik tiriant konkrečius žodžių vartosenos atvejus, atpažinti leksines samplaikas ar sudurtinius žodžius, patikrinti, kuris rašybos variantas labiau paplitęs dabartinėje kalboje ir pan.

Šiam tyrimui surinktos 295 vokiečių kalbos žodyne *Duden. Deutsches Universalwörterbuch* (2007) fiksuotos verbonominalinės konstrukcijos ir sukurtas 120 tūkst. žodžių bandomasis tekstynas. Tekstynui sudaryti reikalingi tekstai rinkti 2010 m. rugsėjo – gruodžio mėn. iš internetinių šaltinių: Vokietijos dienraščio DIE WELT, interneto svetainių FOCUS ir ECHO bei savaitraščių SPIEGEL ir ZEIT. Kadangi šiam tyrimui nebuvo svarbus konkretus teksto žanras, straipsnio tematika ar jo apimtis, tekstai atrinkti remiantis atsitiktinumo principu. Publicistiką nutarta pasirinkti todėl, kad šio žanro

tekstuose verbonominalinių konstrukcijų pasitaiko dažniausiai, o internetinė spauda tyrimo autorėms buvo paranki dėl elektroninio formato.

Išanalizavus visas tiriamąsias konstrukcijas nutarta pagal struktūrą jas skirstyti į mažesnes grupes, darant prielaidą, kad tos pačios formaliosios struktūros konstrukcijoms identifikuoti užteks vieno algoritmo, o kelis mažesnius algoritmus vėliau pavyks sujungti į vieną. Rastos konstrukcijos pagal jas sudarančius struktūrinius vienetus buvo suskirstytos į 6 grupes:

1. daiktavardis + veiksmažodis (*Hilfe leisten* ‘suteikti pagalbą, padėti’);
2. daiktavardis + veiksmažodis + (prielinksnis) (*Abschied nehmen* ‘atsisveikinti’);
3. žymimasis artikelis + daiktavardis + veiksmažodis (*den Schluss ziehen* ‘daryti išvadą’);
4. nežymimasis artikelis + daiktavardis + veiksmažodis (*einen Beitrag leisten* ‘prisidėti’);
5. prielinksnis + daiktavardis + veiksmažodis (*in Gefahr bringen* ‘įstumti į pavojų’);
6. prielinksnis + artikelis + daiktavardis + veiksmažodis (*zum Entschluss kommen* ‘pasiryžti’).

Kadangi šiame straipsnyje norime pristatyti ne visus sukurtus algoritmus, o pirmiausia aptarti patį algoritmų kūrimo principą, todėl pateiksime ir išsamiau pakomentuosime tik vieną algoritmą. Jis skirtas prielinksnis + daiktavardis + veiksmažodis tipo struktūroms atpažinti.

Peržvelgus iš žodynų surinktas fiksuotas verbonominalines konstrukcijas rasta net 101 anksčiau minėto tipo konstrukcija. Mažiausiai šio tipo junginiuose įvairuoja prielinksnio vartojimas – pasitaiko vos 4 vokiečių kalbos prielinksniai (*in, außer, zu, unter*), kiek daugiau yra skirtingų veiksmažodžių (*bringen, bleiben, gehen, geraten, haben, halten, kommen, nehmen, sein, setzen, stehen, stellen, ziehen*), su kuriais gali jungtis labai daug vokiečių kalbos daiktavardžių.

Kadangi tekstynų analizei naudojami kompiuteriniai programiniai įrankiai, konkretaus tekstyno kalbai aprašyti reikalingų taisyklių rinkinį turime pateikti kompiuteriui suprantama kalba. Kaip jau minėta, straipsnyje pristatoma programa UNITEX vartotojui šiam tikslui siūlo naudotis algoritmų kūrimo funkcija. Norėdami naudotis šia funkcija atliekame visus antrame šio straipsnio skyriuje aprašytus veiksmus. Kadangi mus dominanti verbonominalinių konstrukcijų grupė yra gausi ir heterogeniška, todėl ją

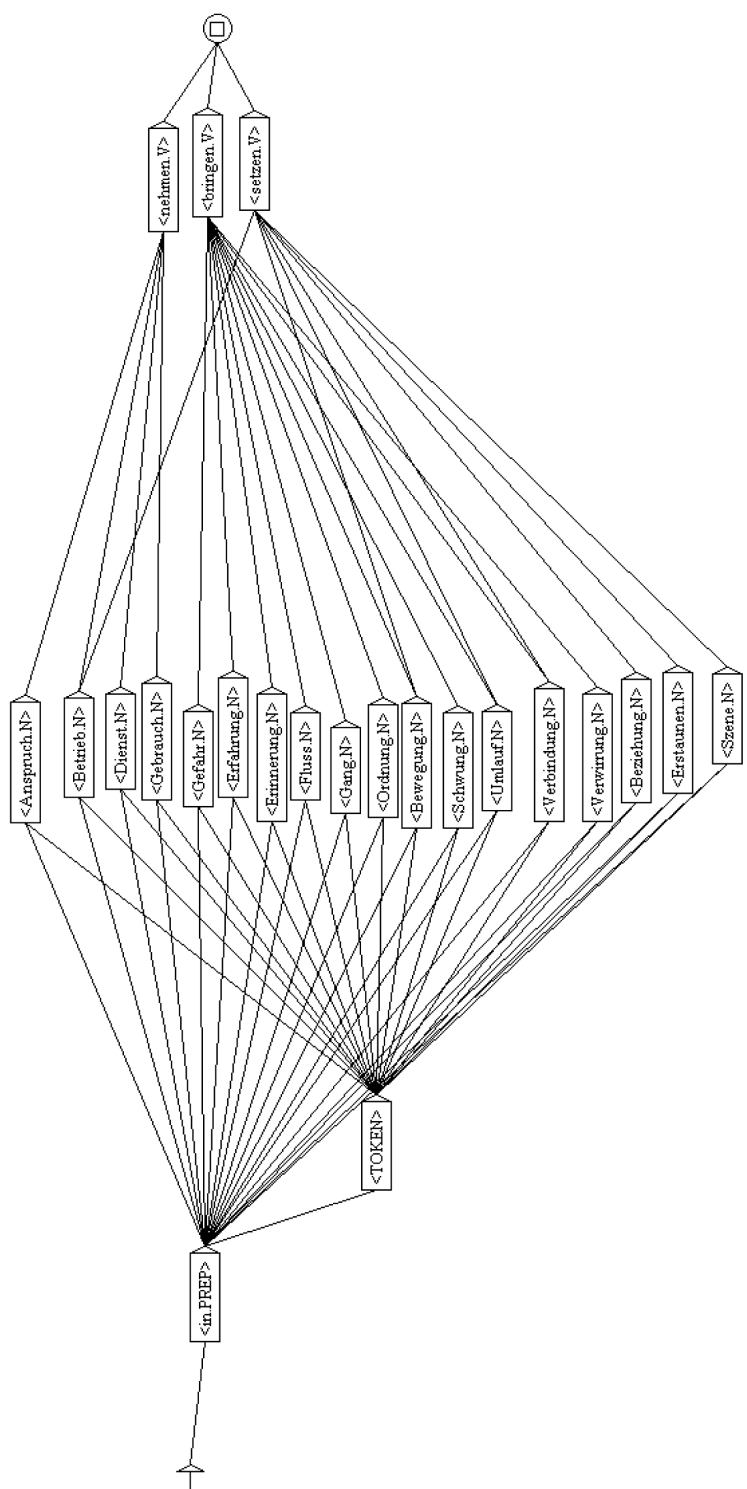
skirstome į pogrupius pagal prielinksni (in, außer, zu, unter). Iš tiriamų verbonominalinių konstrukcijų buvo rastos 23 konstrukcijos su prielinksniu in. Jose vartojami 23 skirtingi daiktavardžiai ir 3 skirtingi veiksmažodžiai. Kadangi algoritmai kuriami ir interpretuojami iš kairės į dešinę, netoli pradžios tašką žyminčios rodyklės, naudodamiesi leksine paieškos kauke, kuriame pirmąjį algoritmo segmentą <in.PREP>. Kiek toliau vieną po kitu išdėstome 23 skirtingus daiktavardžius ir 3 skirtingus veiksmažodžius žyminčius segmentus, pvz., <Anspruch.N> ir <nehmen.V>. Tokia segmentų išdėstymo tvarka pasirinkta neatsitiktinai. Analizuojant bandomąjį tekstyną pastebėta, kad dėl griežtos žodžių tvarkos vokiečių kalbos sakiniuose verbonominalinės konstrukcijos labai dažnai atsiduria sakinio gale, nes yra sudėtinio tarinio dalis, pvz.:

- (1) *Man muss alles in Anspruch nehmen.* (prielinksnis + daiktavardis + veiksmažodis)
'Reikia viskuo pasinaudoti.'

Taip pat pastebėta, kad bandomajame tekстыne pasitaiko atvejų, kai prielinksnio, daiktavardžio ir veiksmažodžio konstrukcijose tarp prielinksnio ir daiktavardžio įsiterpia būdvardis. Kurdami algoritmą į tai taip pat turime atsižvelgti, todėl tarp prielinksnio ir daiktavardžių segmentų sukuriame dar vieną segmentą <TOKEN>. Ši kaukė koduoja bet kokią leksemą.

Išdėliotus algoritmo segmentus turime sujungti. Norėdami sujungti du segmentus, pirmiausiai spragtelime pele ant kairiau esančio, tuomet ant dešiniau esančio segmento, o programa automatiškai nubrėžia linijas. Visos nagrinėjamos konstrukcijos turi prielinksni in, todėl jo segmentą reikia sujungti su visais daiktavardžiais. Jungdami daiktavardžių ir veiksmažodžių segmentus, atsižvelgiame į tai, kad konkretūs daiktavardžiai yra vartojami ne su visais, bet tik su vienu ar dviem veiksmažodžiais, pvz., in *Anspruch nehmen* 'pasinaudoti', bet in *Betrieb nehmen* 'pradėti eksploatuoti' arba in *Betrieb setzen* 'pradėti eksploatuoti'. Kitas algoritmo kelias veda nuo prielinksnio in bet kurią leksemą žyminčio segmento link, toliau prie daiktavardžio ir veiksmažodžio. Sujungę visus segmentus, gauname 9 pav. pavaizduotą paieškos algoritmą ir jį išsaugome.

Norėdami įsitikinti, ar mūsų algoritmas atpažįsta konkrečias struktūras, programoje UNITEX renkamės meniu punktą *Text → Locate Pattern...* ir užklauso lange nurodome, kad paiešką atliksime pagal mūsų sukurtą algoritmą (meniu punktas *Graph*). Paieškos rezultatų lange programa informuoja, kad naudojant 9 pav. pateiktą algoritmą, bandomajame tekстыne atpažinta 30 prielinksnis + daiktavardis + veiksmažodis struktūros konstrukcijų. Nurodžius programai sudaryti konkordansą, gauname 10 pav. pateiktus rezultatus.



9 pav. Paieškos algoritmas prielinksnis + daiktavardis + veiksmazodis tipo struktūroms atpažinti

Šis sukurtas algoritmas labai mažas. 295 žodynuose fiksuotas verbonominalines konstrukcijas pagal struktūrą straipsnio autorėms pavyko suskirstyti į 6 grupes ir 27 pogrupius ir kiekvienam pogrupiui sukurti po algoritmą. Iš viso programa UNITEX bandomajame tekстыne atpažino 274 verbonominalinių konstrukcijų vartojimo atvejus. Pabandžius sujungti kelis mažesnius algoritmus, kartais viskas vyko sklandžiai, tačiau pasitaikė ir tokių atvejų, kai kilo problemų: programa atpažindavo ne visiems pogrupiams priklausančias konstrukcijas, aptikdavo tik dalį anksčiau rastų junginių arba retais atvejais visai neveikė. Taigi tikslas – sukurti vieną algoritmą vokiečių kalbos verbonominalinėms konstrukcijoms atpažinti – kol kas lieka neįgyvendintas, tačiau tikime, kad tai tėra laiko klausimas ir uždavinys ateičiai.

ermieter müssen Haftpflichtversicherung in Anspruch nehmen{S} Auch wenn Mieter Schäden verursac
sofern dies erforderlich ist.Leistungen in Anspruch nehmen. ggf staatliche Zuschüsse - Hartz IV
rm wegen Portugal.Wer kann WLAN-Support in Anspruch nehmen?{S} Studierenden und Mitarbeitern de
abgeschlossene Haftpflichtversicherung in Anspruch nimmt.{S} Dieser Anspruch folgt daraus, das
mieter hat, dass dieser den Versicherer in Anspruch nimmt.{S} Tut er das nicht, kann das nicht
nden gestrichen.{S} Einen neuen Drucker in Betrieb nehmen | schieb.de - [Išversti ši puslapį]
skutinis įrašas: 18 vasario{S} Flugzeug in Betrieb setzen.{S} Guten Abend, leider konnte ich es
ein Flugzeug ...{S} Verbotsschild Nicht in Betrieb setzen (Text) Verbotsschild nach DIN-4844,
Ausstieg hat die Atomkraftgegner wieder in Bewegung gebracht.{S} SPD und Grüne rücken zusammen,
anderthalb Wochen seither haben vieles in Bewegung gesetzt.{S} Im Kopf, vor allem.{S} Es ist n
ar Hammes-Lerch jetzt an.Das Schwungrad in Bewegung setzen{S} Der 27. Parteitag der KPD/SU - Sta
racht. ...{S} Wie ehemalige Schuldtitel in Erfahrung bringen?Kann ich meine genetische Prädispo
enetische Prädisposition für Brustkrebs in Erfahrung bringen? Jede 8. Frau erkrankt im Laufe i
eit und Anmut, die jedes denkende Gemüt in Erstaunen setzen. ..12 žin. - 4 autor. - Paskutinis
elektrischen Impuls{S} das Herz wieder in Gang bringen.Bereits bei leichten Rhythmusstörungen,
hrrad gegen den Willen des Berechtigten in Gebrauch nimmt, wird mit Freiheitsstrafe bis zu drei
m er die Sache nicht wie ein Eigentümer in Gebrauch nimmt und alles unterlässt, ...{S} Autofahr
, ...{S} Autofahren...{S} Andere Leute in Gefahr bringen.{S} Griechen bringen Euro in Gefahr{S
wenn man sein eigenes Leben dabei nicht in Gefahr bringt.{S} Wer nicht hilft wird der unterlass
dem auch die Wikinger folgten.{S} Sieg in Gefahr gebracht{S} zuletzt aktualisiert: 12.10.2006
hrungen sollen Stadtzentrum von Staufen in Gefahr gebracht haben.Raucher bringt Familie in Gefa
der vielleicht doch nicht?{S} Sanitärer in Gefahr gebracht - Würzburg | Nachrichten - mainpost.
lenburg) sich und drei weitere Menschen in große Gefahr gebracht. ...{S} Raucher bringt Familie
lenburg) sich und drei weitere Menschen in große Gefahr gebracht. ...{S} Raucher bringt Familie
lenburg) sich und drei weitere Menschen in große Gefahr gebracht. ...{S} Wie ehemalige Schuldti
ig.{S} Die Kanzlerin müsse "ihren Laden in Ordnung bringen", sagte Steinmeier.{S} Nach kaum sec
ialdemokraten werden die Dinge.Die Welt in Ordnung bringen.{S} Ein kleiner Junge kam zu seinem
U und damit auf die lange Bank.{S} Haus in Umlauf gebracht{S} Bild: Bruno Arnold{S} Die Börsen
2010 den Steintaler als lokale Währung in Umlauf bringen.{S} Das kündigte Vereinschef Lothar H
ene Formate.{S} Mit wem sollten wir uns in Verbindung setzen, wenn wir uns selber einen Aufzug

10 pav. Identifikuotų prielinksnis + daiktavardis + veiksmožodis tipo struktūrų konkordansas

5 Baigiamosios pastabos

Šiame straipsnyje pristatytos svarbiausios programos UNITEX funkcijos: tekstynų ir paieškos algoritmų kūrimas, dažninių žodžių sąrašų generavimas ir konkordansų gavimas. Nors šia programa kol kas negalima apdoroti lietuvių kalbos tekstų, autorių manymu, ateityje ji galėtų gerokai palengvinti ir paspartinti Lietuvos tyrėjų darbą. Jei ši

įrankių pavyktų pritaikyti lietuvių kalbai, jis padėtų kurti naujus kalbos analizei skirtus įrankius, internetinės paieškos sistemas, mašininio vertimo programas ar rašybos ir gramatikos tikrintuvus.

Antroje straipsnio dalyje pristatyta viena šios programos praktinio taikymo galimybių – autorių sukurtas vokiečių kalbos verbonominalinių konstrukcijų paieškos algoritmas. Tirti verbonominalines konstrukcijas pasirinkta dėl aiškios ir stabilios jų struktūros bei dažno vartojimo vokiečių kalboje. Nors autorėms iš viso pavyko sukurti 27 algoritmus, šiame straipsnyje pristatomas vos vienas, kadangi norėta išanalizuoti ne atskirus algoritmus, o aptarti patį jų kūrimo principą. Tenka pripažinti, kad tikslas – sukurti vieną algoritmą vokiečių kalbos verbonominalinėms konstrukcijoms atpažinti – kol kas lieka neįgyvendintas, nes bandant sujungti atskiras dalis į vientisą darinį susidurta su tam tikromis problemomis, tačiau autorės tikisi, kad tai tik laiko klausimas.

Duomenų šaltiniai

- Duden *Deutsches Universalwörterbuch*. 2007. Hrsg. von der Dudenredaktion. 6., überarb. und erw. Aufl. Mannheim, Leipzig, Wien, Zürich: Dudenverlag.
- Die Welt* Interneto prieiga: <http://www.welt.de/>, žiūrėta 2010-12-15.
- Focus* Interneto prieiga: <http://www.focus.de>, žiūrėta 2010-12-15.
- Echo* Interneto prieiga: <http://www.echo-online.de/>, žiūrėta 2010-12-15.
- Spiegel* Interneto prieiga: <http://www.spiegel.de/>, žiūrėta 2010-12-15.
- Zeit* Interneto prieiga: <http://www.zeit.de/index>, žiūrėta 2010-12-15.

Literatūros sąrašas

- Alosevičienė, E. 2004. Junginiai su funkciniais veiksmažodžiais: vertimo problemos. *Tekstas ir kontekstas: prasmės formavimasis*. Kaunas: VU KHF. 127–135.
- Daniels, K. 1963. *Substantivierungstendenzen in der deutschen Gegenwartssprache. Nominaler Ausbau des verbalen Denkkreises*. Düsseldorf: Schwann.
- Engelen, B. 1968. Zum System der Funktionsverbgefüge. *Wirkendes Wort* 16, 289–303.
- Grigonytė, G., E. Rimkutė 2005. Automatinis lietuvių kalbos veiksmažodžių grupių atpažinimas. *Konferencijos „Informacinės technologijos 2005“ pranešimų medžiaga*. Kaunas. 315–320.
- Gross, M. 1997. The Construction of Local Grammars. E. Roche and Y. Schabes (eds.). *Finite-State Language Processing*. Cambridge: The MIT Press. 329–354.

- Helbig, G., J. Buscha. 2001. *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Leipzig, Berlin und München: Langenscheidt.
- Hentschel, E., H. Weydt. 1990. *Handbuch der deutschen Grammatik*. Berlin, New York: Walter de Gruyter.
- Heringer, H. J. 1968. Die Opposition von „kommen“ und „bringen“ als Funktionsverben. *Sprache der Gegenwart* 3. Düsseldorf: Schwann.
- Kniūkšta, P. 2005. *Administracinė kalba ir jos vartoseną*. Vilnius: Lietuvių kalbos institutas.
- Körösi, V. 2008. *Datenbank zur Analyse deutscher Sätze an der Grenze zwischen regulären Stützverbkonstruktionen und idiomatischen Satzrahmen*. Dissertationsschrift. München: LMU.
- Marcinkevičienė, R. 2000. Tekstynų lingvistika. Teorija ir praktika. *Darbai ir Dienos* 24, 7–64.
- Mel'čuk, F. 2004. *Automatic Recognition of Organization Names in English Business News*. PhD thesis. Ludwig-Maximilians-Universität München.
- Paumier, S. 2008. *Unitex 2.0. User Manual*. Interneto prieiga: <http://www-igm.univ-mlv.fr/~unitex/UnitexManual2.1.pdf>, žiūrėta: 2011-11-15.
- Pottelberge, J. van. 2001. *Verbonominale Konstruktionen: vom Sinn und Unsinn eines Untersuchungsgegenstandes*. Heidelberg: Winter.
- Utkā, A. 2000. Kalbinė įranga ir jos galimybės. *Darbai ir Dienos* 24, 275–285.

Das Programm UNITEX und seine Anwendungsmöglichkeiten

Skaistė Volungevičienė, Alina Isilionis

Zusammenfassung

In den letzten Jahrzehnten ist der Computer zu einem unentbehrlichen Teil des linguistischen Alltags geworden. Man kann sich linguistische Untersuchungen ohne Computer und verschiedene Tools kaum noch vorstellen. Dies gilt auch für Litauen, obwohl die Korpuslinguistik in unserem Land noch keine so lange Tradition hat: für das Litauische sind bisher das Rechtschreibprogramm *Juodos avys*, das Silbentrennungstool *Skie-muo* und das Tool für die morphologische Analyse *Lemuoklis* entwickelt worden. Auf unserem Markt beobachtet man aber eher die Tendenz, nicht bereits entwickelte Produkte eigenen, neuen Zwecken anzupassen, sondern neue zu entwickeln. Deswegen erscheint es sinnvoll, das Programm UNITEX, mit dem man sowohl die lexikalische, die morphologische als auch die syntaktische Analyse der Korpora durchführen kann (und dies sogar in 17 Sprachen), vorzustellen. In diesem Beitrag werden die wichtigsten

Funktionen des UNITEX näher beschrieben und ein konkretes Anwendungsbeispiel, nämlich ein Graph für die Erkennung der Funktionsverbgefüge im Beispielkorpus, vorgestellt. Der Beitrag ist keinesfalls als eine ausführlichere Benutzungsanweisung zu verstehen, sondern soll vor allem das Interesse der litauischen LinguistInnen an dem Programm wecken und der Bekanntmachung mit seinen Funktionen dienen. Die Autorinnen würden sich freuen, wenn es in Zukunft möglich wäre, das Programm UNITEX auch für die Analyse litauischer Korpora anzuwenden.

[teikta 2010 m. lapkričio mėn.