CoRoLa
LANSATĂ LA BUCUREȘTI ȘI LA CHIȘINĂU

Doctor habilitat în filologie **Elena UNGUREANU**
Institutul de Filologie
Institutul de Dezvoltare a Societății Informaționale

La Chișinău, în data de 25 mai a. c., la Academia de Științe a Moldovei, a fost lansat proiectul prin care a fost creat *Corpusul computațional de referință pentru limba română contemporană*, denumit CoRoLa (COrpus of ROmanian LAnguage, http://corola.racai.ro/), realizat în perioada 2014–2017, proiect prioritar al Academiei Române. La elaborarea acestuia și-au adus contribuția Institutul de Cercetări pentru Inteligență Artificială „Mihai Drăgănescu" din București și Institutul de Informatică Teoretică din Iași. La eveniment au participat cercetători de la Institutul de Filologie, de la Institutul de Matematică și Informatică și de la Institutul de Dezvoltare a Societății Informaționale.

CoRoLa reprezintă o colecție de texte contemporane (de după 1945), în versiune scrisă și orală, având dimensiuni foarte mari (circa 1 miliard de ocurențe, adică întrebuințări ale cuvintelor, și peste 300 de ore de înregistrări vocale). Corpusul este însoțit de un set de metadate care se referă la autor, data publicării, editură, genul literar al textului etc., precum și de numeroase adnotări care constituie informații de natură lingvistică și gramaticală.

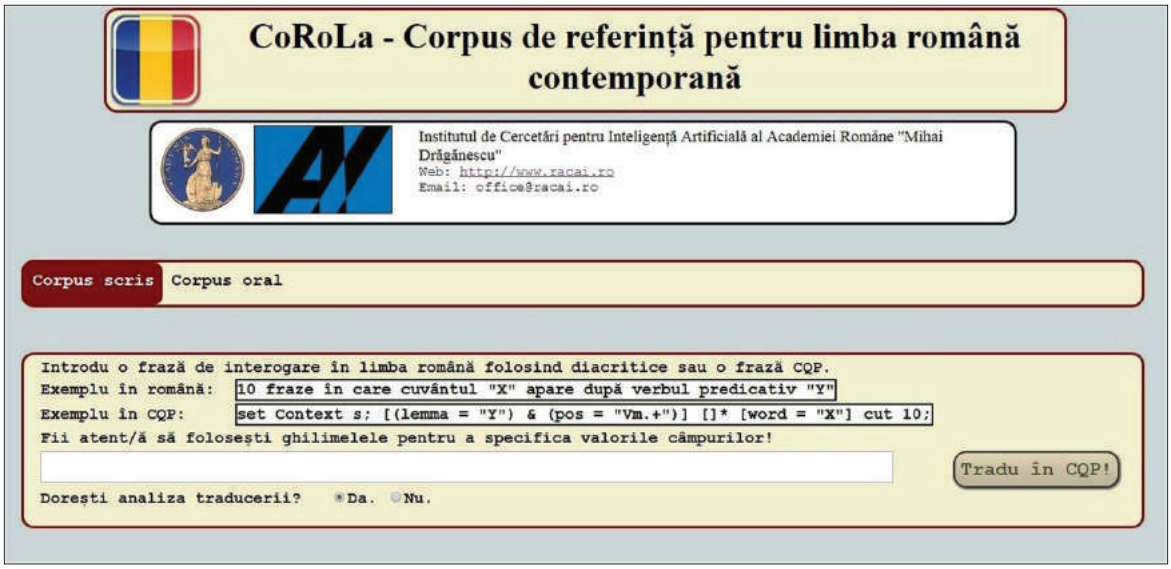
Evenimentul s-a desfășurat sub genericul „**INFORMATICA ÎN SLUJBA LIMBII ROMÂNE. *Resurse și cercetări de lingvistică computațională*"** și a adus în vizorul publicului interesat problematica foarte actuală a corpusurilor computaționale, cel mai important produs de acest gen în spațiul limbii române fiind CoRoLa. O echipă formată din cercetători și doctoranzi de la Institutul de Informatică Teoretică, Filiala Iași a Academiei Române, au prezentat următoarele comunicări: Dan CRISTEA. *CoRoLa, sau petalele informatice ale limbii române* (O privire cuprinzătoare asupra resurselor de lexicografie computațională românească existente și în dezvoltare, precum și o viziune integratoare asupra lor); Mihaela ONOFREI. *CoRoLa în mâinile lingviștilor* (Exerciții de acces la corpus); Cecilia BOLEA. *Ajutați-ne să creștem CoRoLa!* (Exerciții de dezvoltare a corpusului adresate voluntarilor); Alex MORUZ. *Noul eDTLR* (Noi dezvoltări la „Dicționarul tezaur al limbii române în format electronic"). Subsemnata împreună cu Igor COJOCARU (Institutul de Dezvoltare a Societății Informaționale) au relevat importanța conceptului de „deschidere" și a ceea ce numim corpus deschis pentru știința deschisă.

Asistența a adresat numeroase întrebări, arătându-și interesul pentru un instrument foarte laborios, care oferă o amplă și prețioasă informație, structurată pe o platformă electronică, în scopul studierii și învățării limbii române.

Echipa a adresat un îndemn tuturor celor interesați de a se implica în calitate de voluntari în vederea completării progresive a proiectului, care va avea un mare impact asupra societății. IDSI a asigurat transmisiunea online a evenimentului.

Prima lansare a proiectului a avut loc în Amfiteatrul „Ion Heliade Rădulescu" al Bibliotecii Academiei Române din București pe 14 decembrie 2017. La eveniment au participat colaboratorii nemijlociți ai proiectului și colaboratori de la Institutul Limbii Germane din Mannheim, considerat creatorul celui mai mare corpus al limbii germane, împreună cu care se așteaptă includerea corpusului CoRoLa într-un circuit multilingual. Proiectul corpusului este susținut de Fundația Alexander von Humboldt. În deschiderea evenimentului au vorbit: acad. Bogdan Simionescu, vicepreședintele Academiei Române; acad. Florin Filip, președintele Secției de Știința și Tehnologia Informației a Academiei Române; dr. Thomas Hesse, Secretar General adjunct al Fundației Alexander von Humboldt, cu un mesaj de salut din partea prof. dr. dr. h. c. mult. Ludwig M. Eichinger, director al Institutului Limbii Germane din Mannheim.

Prezentarea proiectului și moderarea evenimentului a revenit acad. Dan Tufiș, director al Institutului de Cercetări în Inteligență Artificială „Mihai Drăgănescu" al Academiei Române, București. În cadrul sesiunii *CoRoLa în context multilingv și multimedia*, conf. univ. dr. Ruxandra Cosma de la Universitatea din București a prezentat comunicarea *CoRoLa in a multilingual context*. Dr. Marc Kupietz, de la Institutul Limbii Germane din Mannheim, a trecut în revistă realizările a trei proiecte – DRuKoLA, KorAP și EuReCo –, invocând, în contextul globalizării, importanța și contribuția lor la dezvoltarea unei infrastruc-



turi comune multilingve pentru cercetarea bazată pe corpus(uri). Directorul Institutului de Informatică Teoretică al Academiei Române (Iași), acad. Horia-Nicolai Teodorescu, a vorbit despre CoRoLa în context multdisciplinar.

De specificat că **DRuKoLa**, finanțat de Fundația Alexander von Humboldt, început în ianuarie 2016 ca o colaborare între Universitatea din București, Institutul Limbii Germane și institutele de cercetare ale Academiei Române în București și Iași, este un proiect transdisciplinar care implică lingvistica corpusurilor, lingvistică computațională, lingvistică aplicată și studii interlingvistice, aplicații informatice și dezvoltarea infrastructurii de cercetare. Obiectivul principal al proiectului constă în construirea, furnizarea și armonizarea corpusurilor comparabile în cele două limbi: germană și română. **KorAP** e o platformă de analiză a Corpusurilor, dezvoltată la Institutul Limbii Germane din Mannheim, lansată în 2011, cu o arhitectură extensibilă și scalabilă. Aceasta e capabilă să administreze și să analizeze cantități foarte mari de date, atât primare, cât și adnotate, oferind utilizatorului acces la ambele versiuni ale textelor. **EuReCo** e un Corpus European de Referință, constituit ca bază durabilă pentru cercetarea interlingvistică, proiect comun care vizează armonizarea a trei corporații europene, DeReKo, CoRoLa și Hungarian National Corpus, în ceea ce privește metadatele, convențiile de adnotare și interfețele de interogare.

Deosebit de utile s-au dovedit a fi demonstrațiile practice de lucru pe materialul corpusului, aspect căruia merită să i se dedice și câteva ateliere. Dr. Verginica Barbu Mititelu de la Institutul de Cercetări în Inteligență Artificială „Mihai Drăgănescu” al Academiei Române (București) a demonstrat audienței cele mai importante *Modalități de interogare a componentei scrise a corpusului*, în timp ce dr. Radu Ion, de la același institut, s-a referit atât la *Interfața de interogare a componentei scrise, cât și a celei orale*.

Conf. univ. dr. Adina Dragomirescu, director al Institutului de Lingvistică „Iorgu Iordan – Al. Rosetti” al Academiei Române (București), în comunicarea *Gramatica și corpusul* a argumentat importanța corpusului pentru descrieri pertinente ale nivelului gramatical al limbii; prof. univ. dr. Rodica Zafiu, șef al Catedrei de Limba Română, Facultatea de Litere, Universitatea din București, s-a referit la tendințele privind frecvența unor termeni și îmbinări de cuvinte și regăsirea acestora cu tot cu contexte în corpusul de referință; dr. Gabriela Haja, de la Institutul de Filologie Română „Alexandru Philippide” al Academiei Române (Iași), a făcut *O confesiune a lexicografului: slăbiciunea sa în fața cuvintelor, a trecerii vremii și a Dicționarului limbii române* (DLR). Subsemnata, reprezentând Institutul de Filologie al AȘM și Institutul de Dezvoltare a Societății Informaționale (Chișinău, Republica Moldova), a propus audienței un *Scurt exercițiu comparativ între două corpusuri: CoRoLa vs IntraText.* La finele evenimentului, prof. dr. Dan Cristea, membru corespondent al Academiei Române, de la Universitatea „Al. I. Cuza” din Iași și Institutul de Informatică Teoretică, s-a întrebat: *Ce e de făcut mai departe?*, răspunsul fiind completarea corpusului în flux continuu, atragerea și instruirea voluntarilor, perfecționarea instrumentelor elaborate și oferirea de servicii informatizate.

Discuțiile care au urmat au arătat importanța unui asemenea proiect de anvergură, nu doar pentru lingviști, ci și pentru publicul larg, dar și problemele care urmează să fie soluționate, revenirea la subiectul privind reprezentativitatea acestuia, precum și nevoia de colaborare, inclusiv cu institutele de profil din Republica Moldova.