***Anexa 12.3 - RAPORT ŞTIINŢIFIC SINTETIC AL PROIECTULUI*** *Extracţie şi gestiune terminologica prin manipularea unui corpus multilingv de texte specializate (domeniul tehnic)-Termteh- Contract nr. înregistrare UPIT 1855/07.02.2019*

**1. Obiectivele proiectului**

Proiectul de cercetare *Extracţie şi gestiune terminologica prin manipularea unui corpus multilingv de texte specializate (domeniul tehnic)-Termteh* a avut ca obiectiv principalconstituirea unei baze de date terminologice care să cuprindă cuvinte din domeniul tehnic, mai precis din domeniul automobilului.

**2. Descrierea succintă a modului şi a gradului de îndeplinire a obiectivelor propuse**

(2 – 4 pag.)

Baza de date terminologice *TradTeh* a fost realizată pornind de la un corpus multilingv (limbile engleză, franceză, germană, română), corpus care cuprinde texte de specialitate avand ca temă Automobilul (componente, procese, operaţii, mecanică, automatizare, dezvoltare durabilă).

Această bază de date terminologice va fi folosită atât în cadrul activitatea de cercetare cât şi în activitatea didactică desfăşurată în cadrul Centrului de limbi Straine Logos şi a Departamentului de Limbi Străine Aplicate.

- Activitatea de cercetare. Este nevoie de prezenţa unui corpus multilingv de texte din domeniul tehnic unde limba română să fie prezentă, care să permită identificarea şi analiza unor problematici de lingvistică aplicată, de traducere specializată, terminologie sau sociolingvistică.

- Activitatea didactică. Ne desfăşuram activitatea în cadrul Departamentului de Limbi străine aplicate, care gestionează două programe de studii: un program de licenţă, *Limbi moderne aplicate* (LMA), şi un program de master interdisciplinar, *Limbaje specializate şi traducere asistată de calculator* (LSTAC). In planul de învăţământ al acestor două specializari există o serie de discipline unde componenta *terminologie* este centrală: cursurile practice de traducere specializată, cursuri de limbaje de specialitate, cursul de Introducere în terminologie, cursul de Politică, planificare şi amenajare lingvistică, cursul de Introducere în lingvistica corpusului. Instrumente de manipulare a corpusurilor şi extracţie terminologică.

In cadrul activităţilor practice, folosim memorii de traducere asistată (TRADOS SDL), programe de extracţie terminologică (AntConc, LancsBox), precum şi programe de stocare şi gestiune terminologică. SDL Multiterm este un program de stocare şi gestiune terminologică fără de care memoria de traducere asistată Trados SDL nu poate funcţiona în parametri optimi (nu furnizează variante de traducere dacă nu are o bază de date terminologice ataşată proiectului de traducere). Baza de date terminologice *TermTe*h va fi folosită în cadrul activităţilor didactice legate de cursurile mai sus mentionate cât şi în activitatea de elaborare a lucrărilor de finalizare studii ale studentilor.

Corpusul alcătuit din resurse electronice viabile din punct de vedere ştiinţific a fost anotat prin preprocesarea manuală asistată de calculator şi prelucrat atât prin procesare automată cât şi prin analize calitative legate de rezultatele oferite de programele informatizate legate de frecvenţa de apariţie, variaţia şi concordanţa lingvistică. Termenii identificaţi au fost analizaţi, validaţi şi introduşi în baza de date terminologice care cuprinde rubrici menite să ofere informaţtii precise cu privire la contextul de folosire, definiţie, etimologie, termeni echivalenţi în celelalte limbi.

Activităţile proiectului au fost eşalonate pe parcursul a 12 luni, proiectul prevăzând activităţi individuale şi de echipă, precum şi reuniuni de lucru pentru discutarea stadiului proiectului. Rezultatele fac obiectul unui articol ştiinţific care va fi publicat într-o revistă de specialitate şi vor fi utilizate în activitatea didactică a membrilor CLS Logos.

Pentru realizarea obiectivului proiectului, acela de a constitui o baza de date terminologice cu termeni din domeniul tehnic, a fost nevoie, în linii mari, într-o primă etapă, de constituirea unui corpus electronic de texte multilingv (texte în limbile română, engleză, franceză, germană), într-o a doua etapă, de preprocesarea textelor, într-o a treia etapă, de extracţia terminologică şi, în final, de studiul şi tratarea termenilor candidaţi care vor fi inseraţi în baza de date.

Prezentăm mai jos în detaliu fiecare etapă a realizării proiectului:

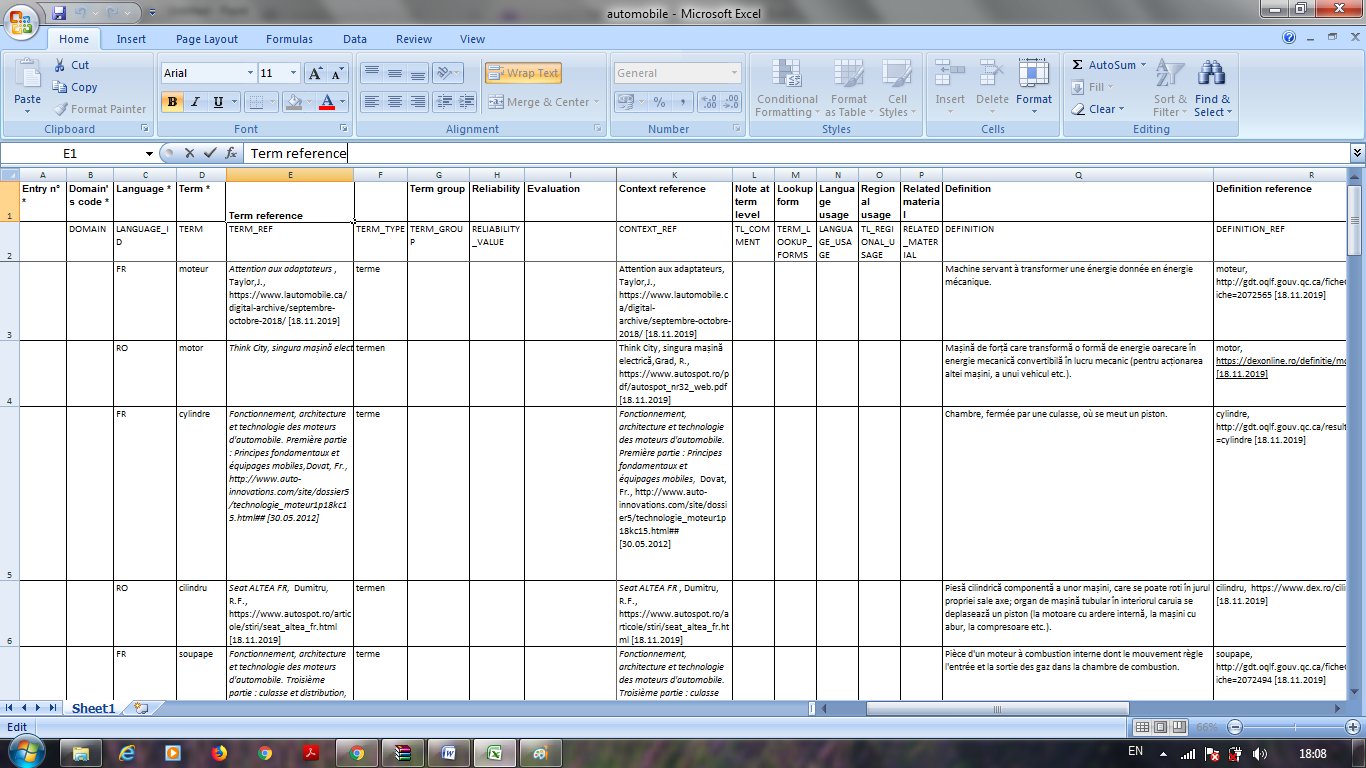
**Etapa1. Constituirea corpusului**. Pentru a restrânge câmpul de investigare pentru acest proiect, am reunit documente scrise legate de automobil: componente, procese, operaţii, mecanică, automatizare, dezvoltare durabilă. Textele reunite sunt în limbile engleză, română, franceză, germană. Pentru fiecare limbă, corpusul este constituit de 50 de texte a cca 7000 de caractere, spaţii incluse. Aceste texte au fost prelucrate prin preprocesare manuală asistată de calculator și procesare complet automată.

**Etapa 2. Anotarea.** Preprocesarea manuală asistată de calculator presupune parcurgerea unor etape menite să uniformizeze textele reunite în corpus atât din punct de vedere grafic cât şi la nivel de conţinut. Este vorba despre procesul de anotare, prin care se adaugă unui corpus informaţie lingvistică interpretativă (Leech, 2005) : lematizare, anotare sintactică, semantică, pragmatică, discursivă, stilistică.

**Etapa 3. Procesarea automată AntConc**. Urmează etapa de procesare automată în programe precum ultima versiune AntConc, dezvoltat de L. Anthony (2018). Au fost urmărite date de frecvenţă, concordanţe lingvistice, de variaţie, cu accent deosebit pe manifestări lingvistice precum colocaţiile.

**Etapa 4. Extracţia şi prelucrarea terminologică**. S-au identificat termenii-candidaţi pentru a putea fi analizati în corpus şi, ulterior, introduşi în baza de date terminologice.

**Etapa 5**. **Constituirea bazei de date terminologice**. Baza de date terminologice (gestionată în document Excel pentru toate cele 4 limbi) conţine rubrici precum: număr înregistrare termen, domeniu, limbă, referinţă termen, context, referinţă context, utilizare regională (daca este cazul), definiţie, referinţe definiţie. S-au stabilit echivalenţe în cele 4 limbi pentru termenii identificaţi.



**Extras bază de date terminologică (domeniul tehnic, limbile franceza-română)**

**Etapa 6. Validarea termenilor**. Următoarea etapă a fost etapa de validare a termenilor retinuţi. Pentru această etapă se va face apel la specialişti cu competenţe în domeniu terminologic şi tehnic din cadrul şi din afara Universităţii din Piteşti.

**Etapa 7. Redactare articol ştiinţific**. Etapa de redactare a unui articol ştiinţific prezintă rezultatele activităţii de cercetare. Acesta este în curs de publicare într-o revistă de specialitate gestionată de CLS Logos.

**Etapa 8**. Ultima perioadă a fost alocată **redactării documentelor de raportare** cu privire la realizarea proiectului.

**3. Modalități de diseminare/valorificare a rezultatelor cercetării**

***3.1. Produse / Tehnologii / Studii / Servicii aplicate***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nr. crt.** | **Denumirea rezultatului** | **Nou / Semnificativ**  **îmbunătățit** | **Numele autorilor** | **Beneficiar** |
| 1 | Baza de date terminologice (engleză, franceză, germană, română), domeniu tehnic | nou | Conf.univ.dr. Cristina Ilinca  Lect.univ.dr.Carmen Bîzu  Lect.univ.dr. Silvia  Boncescu  Student Roberta Cozmîncă | Universitatea din Piteşti |

***3.2. Modele fizice, modele experimentale, modele functionale, prototipuri, normative, proceduri, metodologii, reglementãri şi planuri tehnice noi sau perfecționate***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nr. crt.** | **Denumirea modelului** | **Nou / Semnificativ**  **îmbunătățit** | **Numele autorilor** | **Beneficiar** |
|  |  |  |  |  |

***3.3. Lucrări (articole) ştiinţifice publicate\****

***3.3.1 Articole în reviste cotate ISI Thomson Reuters***

*1.* Autor(i), *Titlu*, detalii

***3.3.2 Articole în volume indexate ISI Proceedings***

*1.* Autor(i), *Titlu*, detalii

***3.3.3 Articole în reviste şi volumele unor manifestări ştiinţifice indexate în alte baze de date internaţionale***

1.**Ilinca**, Cristina., **Bîzu**,Carmen, **Boncescu**, Silvia, **Cozmîncă** Roberta,«Extraction et gestion terminologique par manipulation d’un corpus multilingue de textes spécialisés (domaine technique) »; *Studii de gramatică contrastivă*, nr. 32/2019 , pp. 79-90

Revista *Studii de gramatică contrastivă* este indexata în baze de date internaţionale precum EBSCO, ERIH PLUS, IndexCopernicus, DOAJ- <http://studiidegramaticacontrastiva.info/indexare-internationala/>

***3.3.4 Lucrări (articole) publicate în reviste şi volume de conferinţe cu referenţi***

***(neindexate)***

*1.* Autor (i), *Titlu*, detalii

***3.3.5 Cereri de brevete de invenţie înregistrate***

*1.* Autor(i)., Denumire, detalii

\* se ataşează copii după acestea

**Avizat,**

**Prorector …………….,**

**Verificat, Comisia CCS**

**Întocmit, Director proiect,**

Conf.univ.dr. ILINCA Elena-Cristina