THE EUROPEAN LANGUAGE OF OTHERNESS

FADING AWAY: THE ACQUIS COMMUNAUTAIRE

Violeta Negrea

violeta.negrea@gmail.com

Abstract

The article focuses on the European linguistic policies as a result of the internal communication needs The applied language of the acquis communautaire represent a new stage to the achievement of the legal lingua franca aiming the full mutual understanding and correct application of the common legal measures. The parallel multilingual acquis communautaire corpora has been developed as a resource for language technologies, cross lingual information, multilingual lexicons, meaning disambiguation, etc. aiming the gradual fading away of otherness of the European citizen.

Key Words: *acquis communautaire, translation standards, language standardization techniques, linguistic otherness fading*

Then European Commission's commitment on applied language development aims not only the improvement of the basic language mastery skills, and applied language standardization, but also the gradual fading away of linguistic otherness of the European citizen. The *acquis communautaire* is the first linguistic tool that the European Commission used to approach the sense of otherness to stimulate free movement, work and learning and contribute to the development of jobs and growth and reduce unemployment and increase living standards.

The corpus of the *acquis communautaire* refers to the human rights and obligations of the citizens of the European Union, deriving from the status of membership of their countries to the European Union. *The Community acquis* or *acquis communautaire, /ˈæki: kəˈmju:nətɛər/*[1] called sometimes, for short, the *EU acquis* and often even shortened to simply *acquis* is the body of cumulated legislation, legal acts, and court decisions of the European Union Law. The French origin of the term, *acquis* has the meaning of "that which has been agreed upon", and *communautaire* makes reference to the "the community" of the countries belonging to the European Union. It includes the *treaties, regulations,* and *directives* passed by the European institutions and the judgments of the

---

[1] International Phonetic Association

European Court of Justice. The term is mostly used with reference to the preparations of the candidate countries to join the European Union, as they must adapt, implement and enforce it in their own country so that they become able to manage their legal issues accordingly. The term *acquis,* which is also associated to the Schengen Agreement, commonly referred to as *Schengen acquis,* was first adopted by the World Trade Organization with reference to the General Agreement on Tariffs (GATT) as "*acquis gattien*". The usage of the term was extended to the description of the rules and regulations of the Council of Europe establishing the activity standards in regulating democracy, rule of law and human rights. The organization for Security and Co-operation in Europe (OSCE) and the Organization for Economic Co-operation and Development (OECD) introduced the term in their strategies to implement the rule of law.

With the wave of the two countries joining the European Union in 2007, the *acquis communautaire* consists of parallel legal texts in 24 languages, Bulgarian, Czech, Danish, German, Greek, English, Spanish, Estonian, Finnish, French, Hungarian, Italian, Lithuanian, Latvian, Maltese, Dutch, Polish, Portuguese, Romanian, Slovak, Slovene and Swedish.

Opportunity for Linguistic Research and Translation Standardization

Language research institutions in all the member countries have been involved into linguistic research which include interdisciplinary approaches.

The languages corpora is an opportunity to cross linguistic research taking into consideration the big number of languages involved and the need to translate common issues into rare languages. The parallel multilingual acquis communautaire corpora has been developed as a resource for language technologies, cross lingual information, multilingual lexicons, meaning disambiguation, etc. The linguistic complexity that the size of the corpora reveals is not an opportunity only for the pairing of the major languages, but also for the less common languages, such as Maltese-Finish, or Maltese Romanian. The linguistic and semantic annotation is used to ease the complexity of the translation task by multiplying the text knowledge and classify documents into hierarchical concepts. (Erjavec, T, Ignat, C., Pouliquen, B, Steinberger, R., 529–540). The process of the document translation has been mostly completed and made available in HTML format on the Web.[2]. The ACQUIS corpus has been currently available in 22 languages by the JRC-Acquis version, the only parallel corpus of this kind and size available. The following table shows the distribution of words, characters per language text body, signature, and annex. [3]:

---

[2] http://europa.eu.int/
[3] http://langtech.jrc.it/JRC-Acquis.html

The translation of the *acquis communautaire* is a complex research process that takes into consideration:
- specific aspects of the languages involved in the translation
- the European Union Institutions and the concepts that they promote
- specific formal rules and regulations of the documents drafting

The process involves not only highly specialized translators, but also experts in the *acquis* domains. The EU member states languages are also subject of parallel translations techniques involving the alignment of all possible 231 language pair translation including the *n-to-n* relationship of aligned sentences, which often differs due to the language grammar structure pair.

The *Romanian Style Guide* for the use of Translators of the *acquis communautaire*, issued by the European Institute of Romania establishes a series of detailed conventions for the translation of the Union documents, as they are not only numerous but also complex and highly professional and specific. It provides uniform translation solutions by enlarging the meaning area of specific Romanian words in order to enable them express European concepts that do not exist in Romanian legal language, explanations for certain options of grammar structures, and the main drafting rules for each type of document.

Language ISO Code	Nr. of Texts	Text body			Signature	Annex	Total Nr Words (Text Signature Annex)
		Total Nr. Words	Total Nr Characters	Average Nr. Words	Total Nr Words	Total Nr. Words	
cs	7983	5979261	38479314	749	609441	2100301	8689003
da	7939	6548461	44444011	825	691894	1599456	8839811
de	7914	6576633	47047334	831	571928	1506847	8654608
el	7782	7377316	47715936	948	559487	1628451	9565254
en	7972	7512013	45150120	942	667978	1752545	9932536
es	7809	7964255	48281455	1020	709279	1832745	10506279
et	7944	4925361	38603952	620	439184	1819226	7183771
fi	7735	5134294	43705813	664	565226	118087	6880397

					7	
fr	7862	7812577	45609935	994	673061	1726720	10212358
hu	7489	5391810	40601868	720	539967	1887476	7819253
it	7872	7264126	46792286	923	707467	1704221	9675814
lt	7966	5386359	39936370	676	625365	1948354	7960078
lv	7980	5656335	39290110	709	461736	2011426	8129497
mt	7639	7230538	43919981	947	505324	2288013	10023875
nl	7882	7339465	47699598	931	712255	1710041	9761761
pl	7968	5974605	43160945	750	668248	2070687	8713540
pt	7848	7851904	47225710	1001	648180	1838833	10338917
ro	5792	5122354	33681450	884	402929	4047393	9572676
sk	5278	3911895	26077956	741	413511	1381471	5706877
sl	7984	5989322	37844883	750	573052	2153138	8715512
sv	7731	6472717	42990411	837	560188	1424887	8457792
Average	**7,636**	**6,353,410**	**42,340,925**	**831**	**585,947**	**1,886,338**	**8,825,695**

The rules presented by the Guide are organized on language levels. It presents: orthographic and punctuation problems, morphology and syntax. Then it makes a brief presentation of the European Union, its institutions and working concepts, primary and secondary legislation, structural specific characteristics of the *acquis* documents, their titles and drafting rules, information regarding publication, standard expressions of the legal language used.

Language Standardization Techniques

The need of legal fairness and appropriate communication leads to the necessity of language global standardization which implies the following linguistic techniques:

1. tokenization which identifying words and punctuation in the text;
2. part-of-speech tagging which assigns context-disambiguated word-level syntactic description to the words,
3. lemmatization (or stemming) which assigns the base form to a word

The process of *tokenization* will break the stream of text up into linguistic units of words, phrases, symbols, or other meaningful language pieces that become the input for further processing such as *parsing* or text *mining*. Tokenization is part of the lexical analysis; the pre-stage of the language processing by which the basic language units are to be processed. Any language analysis or generation cannot be carried out without a clear-cut segregation of the basic language units. The identification errors of units that do not need to be further decomposed for subsequent processing are very likely to induce more errors at later stages of text processing.

The following *parsing*[4] stage is based on grammatical division and sentence diagrams[5] that make understanding of the sentence meaning.

The word marking up by the process of *grammatical tagging* or *word-category disambiguation* is based on both definition as well as its context that makes the craft of compiling, writing and editing the specialized language dictionaries. The standardization of the *acquis communautaire* legal terminology is the result of the following applied linguistics considerations:

- identification of the intended users 'needs
- definition of the communicative and cognitive functions of the dictionary
- selection and criteria organizing of the dictionary components
- structure appropriateness for the presentation of the data in the dictionary (i.e. frame structure, distribution structure, macro-structure, micro-structure and cross-reference structure)
- entries systematization
- selection of the collocations criteria structure (adjective+noun, noun+noun (such as collective nouns), verb+noun, adverb+adjective, verbs+prepositional phrase (phrasal verbs), and verb+ adverb
- lemmatization of word or part of word
- definition of words
- definition organization

---

[4]The term comes from Latin *pars* (*orationis*), meaning part of speech
[5] pictorial representation of the grammatical structure of a sentence. Its purpose is to make clear the relations between words and the nature of the syntactic structure to help asses which sentences are and are not possible.

- specifying pronunciations
- labeling definitions and pronunciations for register and dialect, where appropriate
- selection of equivalents
- translation of collocations, phrases and examples
- designing the data access in printed and electronic dictionaries

Lemmatization is the process by which the inflected forms of a word are grouped together so that they can be analyzed as a single item.[6] Specialized dictionaries establish the algorithm of the particular word *lemma* conventionally to represent a lexeme. The process of lemmatization is closely related to stemming. While lemmatization involves context reference, the stemmer operates without contextual meaning difference, reducing accuracy[7]. Lemmatization removes the inflectional ending of the word making the lemma to correspond to the singular form in the case of a noun, the infinitive form in the case of a verb, and the positive form in the case of an adjective or adverb. We can think of lemmatization as a normalization process in which different morphological variants of a word are mapped into the same underlying lemma so they can be analyzed as a single item (term or concept). By reducing the total number of distinct terms, lemmatization decreases the complexity of the analyzed text, and therefore brings important benefits to downstream text processing components. (*Liu, Haibin: Christiansen, Tom; Verspoor Karin: 2012)*,

Conclusions

The policy of applied language standardization is the first step to gradual fading away of linguistic otherness of the European citizen. The *acquis communautaire* is a linguistic tool that made the European Commission approach the sense of otherness of the European Union citizen gradual fading away of linguistic otherness of the European citizen by easing free movement, work and learning and contribute to the development of jobs and growth and reduce unemployment and increasing living standards.

BIBLIOGRAPHY

- Tomaz Erjavec, Camelia Ignat, Bruno Pouliquen: Ralph Steinberger: Massive multi lingual corpus compilation: Acquis Communautaire and

---

[6] Collins English Dictionary, entry for "lemmatize"

[7] Some analyzers store the base stemmed format of the word without meaning, but taking into account the semantics of the word formation only, because the purpose of stemming is to map different forms of a word to a single form, on a rules-based algorithm.

totale, Archives of Control Sciences Volume 15(LI), 2005, No. 4, pages 529–540

- Mathew, Alan, Recreating Europe – The European Union's Policy towards Central and Eastern Europe; Cambridge University Press, 1998
- Acordul interinstituţional din 22 decembrie 1998 privind orientările comune referitoare la calitatea redactării legislaţiei comunitare, JO C073 din 17.031999, pp 0001-0004
- European Commission: Agenda 2000, 16 iulie, 1997
- European Commission: Glossary – Institutions, policies and enlargement of the European Union, Office for Official Publications of the European Communities, 2000
- European Commission: Legislative drafting. A Commission Manual, 2002
- Serviciul de traduceri al Comisiei Europene: English Style Guide, versiunea noiembrie, 2001
- Serviciile Juridice ale Parlamentului European, Consiliul Uniunii Europene şi Comisiei Europene: Practical Guide for persons involved in drafting legislation within the Community institutions, Bruxelles, 2001
- http://eurovoc.europa.eu/drupal/?q=request&uri=http://eurovoc.europa.eu/210682
- http://ipsc.jrc.ec.europa.eu/index.php?id=198