WAYS TO ENHANCE LEGAL DICTIONARIES

Imre Attila, Assoc. Prof., PhD, Sapientia University of Tîrgu Mureș

Abstract: The present paper tries to detect the available legal dictionaries for Romanian and English published in Romania. After their brief presentation, the author will discuss in details two of them, trying to highlight their strengths and weaknesses. However, the selection will not be at random, as we will take into account the number of entries they include, as well as their popularity among users. The article will also argue for the topicality of term bases, which may also represent the next stage of data collection regarding legal terms. The conclusion – on the one hand – will discuss the importance and usability of legal terms from the point of view of computer assisted translation, whereas on the other hand we will try to offer solutions how to enhance the quality of the present-day Romanian-English dictionaries.

Keywords: legal dictionary, term bases, quality, Romanian, English.

Introduction

The importance of translation and interpretation – at least within the European Union – is on the rise (Gambier, 2009). Among the theories, why this is an important and necessary development, globalization and the (r)evolution of technology are two obvious ones (Imre, 2013). As a result, they have had an enormous impact on the (r)evolution of translation, which has led to the fact that modern translators cannot face the market requirements unless they can handle effectively computers and the Internet. Thus they must be familiar with the possibilities and limits of machine translation and computer-assisted translation (often called *translation environments*), as – according to Gouadec – the days of "pencil and rubber" are numbered: "The PRAT or Pencil and Rubber-Assisted Translator is clearly on the way out, though there are still a few specimens at large. The Computer-Assisted Translator has taken over." (Gouadec, 2007, p. 109).

If we take Gouadec's words for granted, then translators should be able to handle two types of databases: term bases (TB) and translation memories (TM), which are two major constituents of all CAT-tools, either separately or combined (*SDL Trados Studio, Déjà Vu, Wordfast, OmegaT, memoQ,* etc.). Of course, handling various formats (document-extensions) leads to the self-evident statement: creating and/or finding these databases is crucial during the work of the modern translator. People usually can handle the Internet rather successfully, some of them can even filter the information very effectively in order to find relevant details about something in particular.

However, we argue that in case of finding large pieces of information (such as the content of large texts, e.g. dictionaries, glossaries), quality is a delicate matter, as even if we are able to find a specific dictionary, we cannot be sure about its quality. Furthermore, the date and circumstances of creating a collection of data (compilation, original, plagiarized) may be relevant

BDD-A22337 © 2014 "Petru Maior" University Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 18:14:09 UTC)

in particular cases, such as dictionaries, glossaries, as languages are in constant change, and due to the evolution of languages, new words and expressions constantly enrich vocabularies. Thus in the next section, we will try to describe our project and explain its topicality.

2. Romanian–English and English–Romanian legal dictionaries in Romania

The initial idea of mapping the available legal dictionaries in Romania stems from the author's contact with courts and tribunals, where translation and interpretation to/from Romanian, English and Hungarian is highly required. During these activities people were often complaining about the poor quality of legal dictionaries in Romania. Professional translators and interpreters who registered on ProZ.com have also mentioned that, so we grew to be interested in the "market" for these dictionaries.

During the search for Romanian–English and English–Romanian law dictionaries we were able to track the following 14 (in alphabetical order):

Nr.	Author(s)	Yea	Title, place, publishing house	Content	Languages
1	Bantaș, A. & Năstăsecu, V.	200	*Dicționar economic român-englez* ești, Niculescu PH	economics,	Ro–En
2	Botezat, O.	201	*Dicționar juridic român-englez / -român,* București, C.H. Beck PH	law	Ro–En, En-
3	Dumitrescu,	200	*Dicționar juridic englez-român* București, Akademos Art PH	law	En-Ro
4	Dumitrescu,	200	*Dicționar juridic român-englez* București, Akademos Art.	law	Ro–En
5	Hanga, V., & , R.	200	*Dicționar juridic englez-român și -englez*. București, Lumina Lex PH	law	En-Ro, Ro–
6	Ionescu- , N.	200	*Dicționar economic englez-român, -englez*. București, Teora PH	economics	En-Ro, Ro–
7	Jidovu, I., & A. & Ghițescu,	201	*Mic dicționar terminologic pentru iul Schengen*. București, Universul PH	economics,	Ro–En–Fr–
8	Lister, R., & K.	201	*Dicționar juridic englez-român / -englez* (trans. By R. Dinulescu) București, Niculescu PH	law	En-Ro, Ro–
9	Lozinschi, S.	200	*Dicționar juridic Român–Englez* București, Smaranda PH	law	Ro–En
10	Mezei, J.	200	*Magyar-román-angol jogi, dasági és üzleti szótár*. București, eck PH	law, nics and ss	Hu–Ro–En
11	Năstăsescu,	200	*Dicționar economic englez-român / -englez*. București, Niculescu PH	economics	En-Ro, Ro–
12	Voiculescu,	200	*Dicționar juridic englez-român / -englez și terminologia UE-SUA* București, Niculescu PH	law	En-Ro, Ro–

BDD-A22337 © 2014 "Petru Maior" University Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 18:14:09 UTC)

13	Voroniuc, A.	199	*Dicționar de termeni economici și i (român-englez)* Iași, Institutul ean PH	economics w	Ro–En
14	Voroniuc, A.	201	*Dicționar englez-român/român-de termeni economici și juridici*. Iași, n PH	economics w	Ro–En, En-

One can easily observe that in some cases we have mixed dictionaries of economics and law, and Ionescu-Cruțan's dictionary is "only" on economics, but it considered to contain many legal terms and expressions. The most known is Hanga and Calciu's dictionary (Hanga & Calciu, 2009), which is at its 5th edition already and the most comprehensible is Lozinschi's (2008). Dictionaries number 7 and 10 have a huge drawback: neither of them contains an index section, thus they are only searchable in Romanian and Hungarian, respectively, which is a problem in the case of dictionaries number 1, 9 and 13 as well.

After having purchased these dictionaries, the major aim was to cross-examine them in an effective way and trying to detect their flaws and correct them. Thanks to a POSDRU project at Petru Maior University in Tg.-Mureș we have embarked upon creating a common database from all these sources. It is our belief that this is possible to carry out within twelve months, due to the fact that these dictionaries should contain the same words and expressions – to a certain extent – belonging to the legal terminology. In case we start with the most comprehensive dictionary, the others are much easier to include in the database. So we considered that we should start with Lozinschi's dictionary, and then compared to the most known one, Hanga and Calciu's fifth edition legal dictionary.

3. Creating the database

As it was mentioned in the introductory part, modern translators should be able to create databases compatible with computer-assisted translation software. The most widespread CAT-tools are compatible with each other – at least, to a certain extent[1]. We have been studying and using *memoQ*[2] and *OmegaT*[3] extensively since 2009, and we can say that creating a term base in *csv* format should be compatible with various, even cross-platform CAT tools. A *csv* format may be easily obtained after having created a *Microsoft Excel* file (*xls* or *xlsx* format) with two columns (in our case Romanian and English), then converted into *csv* format. This type of format can be easily used as an external term base for translation environments. So after collecting the data from dictionary 9, we obtained the following number of entries:

A	6401	G	2942	M	3973	TȚ	3879
B	2339	H	1032	N	3516	U	1131
C	7207	I	6483	O	3116	V	2268
D	4725	Î	7410	PQ	12471	WXY	19
E	3567	JK	1363	R	7284	Z	556
F	3758	L	2978	SȘ	8755	**TOTAL**	**97173**

[1] A list of notable CAT tools may be checked here: http://en.wikipedia.org/wiki/Computer-assisted_translation, 04. 10. 2014.
[2] http://kilgray.com/products/memoq, 05. 10. 2014.
[3] http://www.omegat.org/, 05. 10. 2014.

BDD-A22337 © 2014 "Petru Maior" University Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 18:14:09 UTC)

Table 1. Number of entries from Lozinschi's dictionary

One should note that when a TB is created, one entry means that one word/expression in the source language "equals" one word/expression in the target language, so if we have three translations for *neplăcut*, then we will have three entries: *neplăcut–unpleasant, neplăcut–disagreeable, neplăcut–discomfortable*. If we have further expressions with *neplăcut*, each of them will count as a different entry: *gust neplăcut–unpleasant taste, urmări neplăcute–unpleasant consequences*, etc. (Lozinschi, 2008, p. 389).

In the second stage we collected data from the most widely known dictionary, number 5, with the following number of entries[4]:

A	928	G	100	M	243	TȚ	260
B	258	H	47	N	184	U	59
C	1081	I	449	O	183	V	176
D	545	Î	199	PQ	651	WXY	0
E	274	JK	137	R	728	Z	29
F	293	L	212	SȘ	581	**TOTAL**	**7617**

Table 2. Number of entries from Hanga & Calciu's dictionary

We can observe that there is an enormous difference in number regarding these two dictionaries; basically, Lozinschi's dictionary contains more than 12 times the number of entries to be found in Hanga and Calciu's dictionary. To express it more visually, let us take a look at the clustered column chart below, which contains only the first six letters of the two dictionaries:

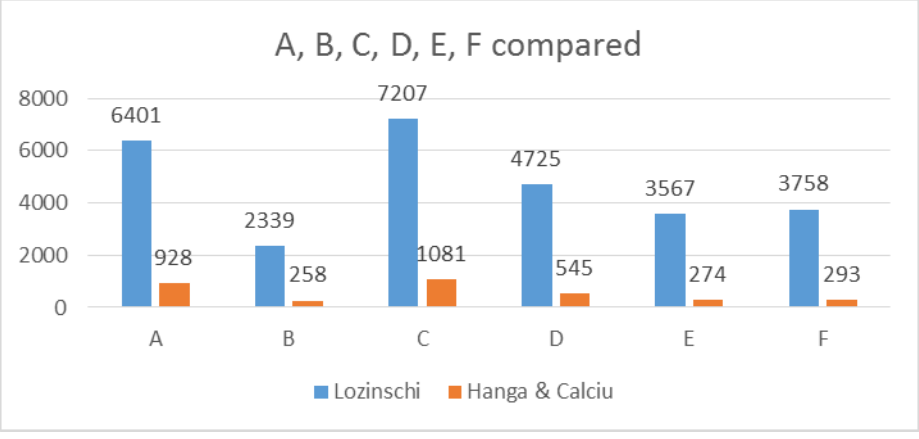


Figure 1. A-F compared

The next step, logically, would be to unify the data from the two sources, but we realized soon enough that important measures for the sake of quality must be taken. While adding the entries to the database, we marked the types of mistakes in the printed version of the dictionaries. Seemingly, three colours are needed; we used red for grievous mistakes, yellow for items to be checked later (entries may be unclear, unjustified, items which – in our humble opinion – do not belong to a legal dictionary, etc.) and green for British English/American English differences

[4] In this stage only the Romanian–English part of the dictionary was accounted for.

BDD-A22337 © 2014 "Petru Maior" University Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 18:14:09 UTC)

(word choice and spelling). We think that this last one is important, as dictionaries either do not bother to clarify which is which (in case both variants are included) or only one version is listed and the author(s) do(es) not explain in the foreword which English is used in it.

The results regarding the quality of these two dictionaries is much worse than imagined before. Hanga and Calciu's dictionary is an extremely widespread one in Romania, very easy to purchase and at low cost. However, professional translators already signalled on proz.com that its quality is below expectations[5]. Lozinschi's dictionary contains fewer errors than the previous one, even if it is much bulkier. We tried to categorize the types of mistakes found, which will be presented and discussed in the next section.

4. Types of errors (Hanga & Calciu's dictionary)

As mentioned in the previous section, there were many errors detected in these two dictionaries. Major types of errors can be the following:

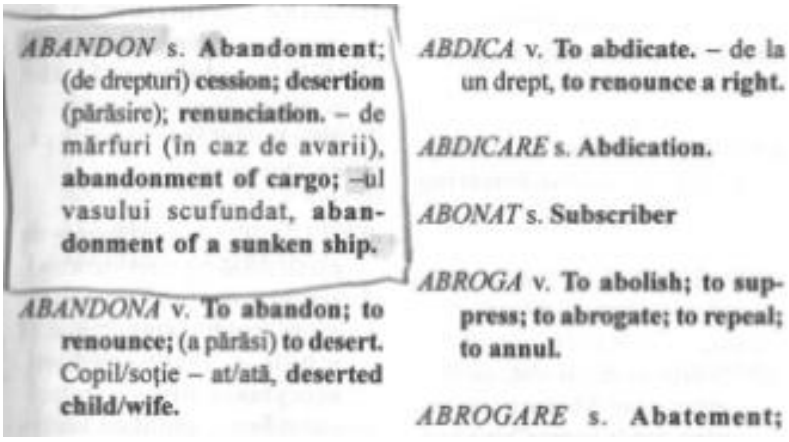
- Translation errors: the translated term is not correct or barely acceptable;
- Unnecessary entries included, which do not belong, in our case, to the legal terminology;
- Formatting mistakes: dictionaries have a standard reference for entry words, translations, symbols, abbreviations, punctuation and layout; deviations from these standards may be either visually bothering or completely wrong;
- Typographical mistakes: even if we know that we can hardly find any printed material without typographical mistakes, there is a reasonable limit above which the reader "feels" that the particular printed material is not acceptable/of poor quality; we tend to believe that this threshold is very low for dictionaries;
- Spelling errors, due to ignorance;
- Grammatical mistakes: the grammatical category of the word is erroneous (e. g. adjective, adverb), but there are other types as well (e. g. negative forms).

After having listed the most typical errors, let us take a look at Hanga and Calciu's dictionary. In our view, the first troublesome thing is the visual effect due to it layout, detailed below:

1. Capitalized and indented main entries: *ABANDON*, *ABANDONA*;
2. First translated word is capitalized, all translations are in bold: **Abandonment; cession**;
3. The symbol for the main entry is not ~ but –;
4. There is a full-stop at the end of the entry, although there are entries without a full-stop: *ABONAT*.
5. Too much "empty" space between the entries.

[5] Here are only two links to prove it: http://www.proz.com/kudoz/romanian_to_english/law_contracts/3490819-prepusi.html, http://www.proz.com/kudoz/english_to_romanian/law_patents/479796-invalidity.html. 04. 10. 2014.

BDD-A22337 © 2014 "Petru Maior" University Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 18:14:09 UTC)


ABANDON s. Abandonment; (de drepturi) cession; desertion (părăsire); renunciation. – de mărfuri (în caz de avarii), abandonment of cargo; –ul vasului scufundat, abandonment of a sunken ship.

ABANDONA v. To abandon; to renounce; (a părăsi) to desert. Copil/soție – at/ată, deserted child/wife.

ABDICA v. To abdicate. – de la un drept, to renounce a right.

ABDICARE s. Abdication.

ABONAT s. Subscriber

ABROGA v. To abolish; to suppress; to abrogate; to repeal; to annul.

ABROGARE s. Abatement;

Figure 2. Sample from Hanga & Calciu, pg. 7

In order to present further problems, in the following we shall mainly confine ourselves to letters I (138 entries) and Î (61 entries), detailed below.

a. **Typographical errors** ("Typographical error," 2014):
   - Mistyped letter: *debit**u**ri* instead of *debit**o**ri* (*IMPOZIT*), **irresponsability* instead of *irresponsibility* (*IRESPONSABILITATE*);
   - Fat-finger syndrome: *rischarging* instead of *discharging*, probably due to the fact that letters 'D' and 'R' are close to each other on the keyboard;
   - Extra letter: **serfdo**r**m* instead of *serfdom* (*IOBĂGIE*);
   - Extra letter due to hyphenation: *pro-operty* (*INALIENABIL*);
   - Missing letter: **indisoluble* instead of *indissoluble* (*INDISOLUBIL*), *loss o civil rights* instead of *loss o**f** civil rights* (*INFAMANT*);
   - Unnecessary hyphenation in the middle of the line: *expen-ses* (*INDEMNIZAȚIE*);
   - Extra string of characters: *travel/travelling* ***ling*** (*INDEMNIZAȚIE*);
   - Words stuck together: ***indubio*** *pro reo* instead of ***in dubio*** *pro reo*.
   - "Atomic typos"[6] are meaningful words in the "wrong" place, thus spellcheckers will not detect them as errors (Bloom, 2012). Our example is *debit**u**ri* ('flow' in plural) instead of *debit**o**ri* ('debtors' under *IMPOZIT*).

In our view, the most grievous types of errors in a dictionary are the ones –whatever type – to be found in the main entry: *INACESIBILITATE* instead of *INAC**C**ESIBILITATE*, or the missing diacritical marks (in our case, Romanian): ***i**nchide* instead of ***î**nchide* (*INTERNA*).

b. **Grammatical errors**: these errors mainly derive from ignorance. In case of verbs, we found cases when the conjugated form was used instead of the infinitive: *închirie**z*** ('I rent', first person singular, present) instead of *închiria* ('to lease', 'to rent'). What is worse, there is a separate entry for *închiria*, so this must have been *închirie**re***, which is a noun ('letting', 'renting'). Further errors include:
   - Singular/plural shift, which remains unmarked: *încasare* ('collection', 'cashing'), whereas an expression rooting from this entry and requiring plural is marked this way: *-i și cheltuieli*, resulting in **încasarei* instead of *încasări*.

[6] http://en.wikipedia.org/wiki/Typographical_error#Atomic_typos, 05. 10. 2014.

BDD-A22337 © 2014 "Petru Maior" University Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 18:14:09 UTC)

- Disagreement in gender between a noun and its modifier (adjective): the main entry is *interlocutori**u*** (adjective, 'interlocutory'), whereas combined with a feminine noun its ending should be *interlocutori**e*** as in *decizie interlocutorie*. Instead, we have *decizie* - in the dictionary, which is understood as *decizie *interlocutoriu*.

c. **Different spelling.** English teachers in Romania usually teach that both UK (British) English and US (American) English spelling are acceptable, but within the same text it should be unified. However, this dictionary is rather inconsistent. In the preface we can read about the (British) English Common Law, and the list of abbreviations contains *SUA*, referring to the United States of America, leading us to the conclusion that British English terms will be used or if there are US variants, they will be marked separately. However, both *judgment* (mainly US) and *judgement* (mainly UK) appear, without specific notice under letters A-I, even if under letter J it is explained: *judgement (judgment SUA*). Further words are: *dishonour, defence, licence, naturalisation*, etc. But the problem is interesting from the point of view of the receptor: who are the target readers? If they are Romanians, then they should see a distinctive sign/abbreviation for the different UK/US spelling. Although it is a minor problem, we have to mention a further nuisance: the dictionary uses three different abbreviations for US spelling: *SUA* (e.g. *ÎNCHISOARE*), *S.U.A.* (e.g. *CĂSĂTORIE*) and *amer.* (e.g. *INTERSTATAL*), which is not really acceptable. In other cases differences in UK/US usage remain unmarked: *jail, gaol, prison, penitentiary, penitenciary* (*ÎNCHISOARE*). However, under *JANDARMERIE* we can find *constabulary* (**în Anglia**), where the explanation in brackets means 'in England', instead of using 'UK'. Anyway, it is a good question whether the translated part (after the main entry) may contain Romanian words or not.

d. **Other errors.** Here we can mention irrelevant entries from the point of view of a legal dictionary (*în orice situație* 'whatever the situation'; *în față* 'in front of'). What is worse, if somebody is looking for 'whatever the situation', then he/she will check it under *orice* 'whatever' or *situație* 'situation', and not under the preposition *în*. Although in few cases, it happens that a Romanian term is used in the translated English: the Romanian *patron* ('employer') is translated as *proprietar* ('owner' in Romanian) instead of *proprietor*. A final, very serious type of mistake we would like to mention (talking about a dictionary) is the wrong alphabetical order under letter C: instead of ***cabi**net, **cabo**taj, **cada**stru, **cadu**c, caducitate*, we have ***cada**stru, **cadu**c, caducitate, **cabi**net, **cabo**taj*. At this stage we think that no further comments are necessary.

5. Ways to enhance dictionaries

It is our firm belief that there are possibilities to enhance a great many of the above presented errors. A thorough check is an option, but it is too late once the dictionary is published.

Typographical mistakes are relatively easy to track if they are retyped in a new office document (*Microsoft Office, Libre Office*) and the spell-checker is set to the desired language. In case both *Microsoft Office* and *Libre Office* are used, we can obtain rather error-free results

BDD-A22337 © 2014 "Petru Maior" University Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 18:14:09 UTC)

(Imre, 2013). This is why we consider it disturbing that Hanga & Calciu's dictionary is at its fifth re-checked and completed edition[7]. A well-founded question is, what is the first edition like?

If we have in mind a term base, then many issues are solved: typo mistakes mainly solved, non-professional layout is solved (term bases have no layout except for the font type and size), similarly to wrong alphabetical order (automatic alphabetical sorting). Furthermore, the missing diacritical marks should not be an issue any more, as they enter the category of typo errors, and if the spelling is set to US English, the great majority of UK English spelling is also signalled and can be labelled as UK/US systematically.

Yet, we will have other problems to solve: when dictionary entries are turned to an electronic database, it is a justifiable expectation to show hits during a search; however, the infinitive verbs forms are not suitable for that, as in texts we are typically faced with conjugated forms, so term bases have two options: they either contain the root (*încurc* instead of *încurca*), or all the possible forms (*încurc, încurci, încurcă, încurcăm, încurcați, încurcat, încurcarăm, încurcaserăm,* etc.), which is a rather long list. Another complicated issue is the conjugated reflexive forms of verbs (***a se*** *întâlni*). A language specific issue is related to *Microsoft or Libre Office*: the Romanian ***ș, ț*** diacritical marks are usually problematic if we use particular font types (e.g. Century Gothic is not so fortunate compared to Times New Roman). This is visible when the *xls* or *xlsx* format is converted to *csv* format for the sake of the database.

Although there are drawbacks of term bases as well, we consider that the gains are far more important, listed below:

- ✓ Extremely large databases may be created ("all-in-one", in our case all 14 dictionaries will be unified into a single one);
- ✓ One entry may contain as many translations as we want to;
- ✓ If the dictionary contains only one direction (e.g. Lozinschi's dictionary), the database may be easily converted into English–Romanian as well, thus no index needed (however absent from Lozinschi's dictionary);
- ✓ Databases are instantly searchable, even if fragments of the words are typed;
- ✓ New entries may be added any time later.
- ✓ Term bases contribute to the systematic quality assurance.

Consequently, we see a far greater future of personally created or downloaded (free or proprietary) specialized term bases, glossaries and translation memories (e.g. DGT Translation Memory, Termium, etc.) and our intention is to create a unified Romanian–English, English–Romanian term base of legal word and expressions until the end of our project in 2015.

The research presented in this paper was supported by the European Social Fund under the responsibility of the Managing Authority for the Sectoral Operational Programme for Human Resources Development (*Sistem integrat de îmbunătăţire a calităţii cercetării doctorale şi postdoctorale din România şi de promovare a rolului ştiinţei în societate*), as part of the grant POSDRU/159/1.5/S/133652.

References

[7] In Romanian: "Ediția a V-a revăzută și adăugită", interior cover/first page.

BDD-A22337 © 2014 "Petru Maior" University Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 18:14:09 UTC)

Bloom, D. (2012, September 30). Spell checkers developing "atomic typo" capabilities. Retrieved October 5, 2014, from http://www.chinapost.com.tw/commentary/the-china-post/special-to-the-china-post/2012/09/30/356026/Spell-checkers.htm

Gambier, Y. (2009). *Competences for professional translators, experts in multilingual and multimedia communication* (p. 7). Brussels: EMT expert group. Retrieved from http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf

Gouadec, D. (2007). *Translation as a Profession*. John Benjamins Publishing.

Hanga, V., & Calciu, R. (2009). *Dicționar juridic englez-român și român-englez*. București: Lumina Lex.

Imre, A. (2013). *Traps of Translation*. Brașov: Editura Universității "Transilvania."

Lozinschi, S. (2008). *Dicționar juridic Român - Englez*. Bucureşti: Editura Smaranda.

Online sources

http://en.wikipedia.org/wiki/Computer-assisted_translation, 04.10.2014.

http://en.wikipedia.org/wiki/Typographical_error#Atomic_typos, 05.10.2014.

http://kilgray.com/products/memoq, 05.10.2014.

http://www.omegat.org/, 05.10.2014.

http://www.proz.com/kudoz/english_to_romanian/law_patents/479796-invalidity.html. 04.10.2014.

http://www.proz.com/kudoz/romanian_to_english/law_contracts/3490819-prepusi.html, 04.10.2014.

BDD-A22337 © 2014 "Petru Maior" University Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 18:14:09 UTC)