Nowadays hardly anyone demands justification for a dictionary digitization project. There seems to be a general agreement on what the digital versions of dictionary can offer to us: much clearer presentation, no space limitations, advanced search capabilities, to name only a few. Practising lexicographers, however, also realize very well how time consuming the completion of such a project can be: instead of advancing their proper work of dictionary making, research teams find themselves sometimes obliged to undertake tasks which distract them from doing «true lexicography». At the same time, prospective advantages do not always seem to compensate for the loss of time spent preparing an electronic dictionary. So it is hardly surprising that some lexicographers, considering the digital version of their dictionaries as in a sense external and only loosely related to their everyday work, still tend to outsource digitization.

Being aware of all the obstacles which such a project might face, in this paper I will argue that dictionary digitization can also be intellectually challenging and may remarkably deepen the professional awareness of the lexicographer, and that for two reasons. First of all, most popular encoding standards necessitate some form of input, in this case, dictionary, standardization. Hence they encourage a thorough analysis of past editorial practices which, in turn, has as its main objective a reduction of «narrative» constituent of lexicographic work as a way to throw into relief linguistic data. Topics of such potential methodological reflection can include:

- different ways of encoding in the paper dictionary the same or similar linguistic information;
- levels of explicitness of linguistic data representation in the paper dictionary;

* This research was partially supported by a research grant of the Polish National Science Centre («eLexicon Mediae et Infimae Latinitatis Polonorum (A-Q)», nr. 3736/B/H03/2011/40).
conflicting classification systems and incoherence of language theories implicit in the paper dictionary as a result of long editorial process.

Secondly, once completed or at least advanced, digital dictionary turns out to be a perfect tool to conduct systematic metalexicographic research.

In the present paper I present the results of the introductory study of the feasibility of automatic term extraction from the *Lexicon Mediae et Infimae Latinitatis Polonorum* (vol. A-Q), dictionary of medieval Latin from Polish sources which has been compiled from 1957 up to the present day. OCR-ized, but not yet proofread text of the *LMILP* has been examined to answer following questions:

a) how are technical terms represented in the *LMILP*?

b) what does one learn about the medieval Latin terminology system from the *LMILP*? What preconceptions of the medieval terminological system can be traced in the *LMILP*?

c) what are the problems that paper dictionary form and its editorial conventions may pose for automatic term extraction?

**Technical Term Marking in the LMILP**

In the *LMILP*, terms are explicitly declared. The most obvious mark of them is the use of *t. t.* label which stands for *terminus technicus*, «technical term». Yet, it never occurs separately, but is always accompanied by subject field label, e. g. *phil. t. t.* «philosophical technical term». This combination, however, is by no means the only possible way of term marking, since special use can be also demonstrated by use of sole subject field label. Suffice to say that on average only 20% of subject field labels are accompanied by *t. t.* Moreover, similar function can be attributed to the expression of type *in textibus* + subject-field description, e. g. *in textibus philosophicis*. This feature has been introduced for the first time in the third volume of the *LMILP* and is now represented by about 200 occurrences.


2 Nancy Ide, Jean Véronis, «Knowledge Extraction from Machine-readable Dictionaries: An Evaluation», in *Machine Translation and the Lexicon*, 1995, p. 17-34, express doubts as to whether machine-readable dictionaries can be a valuable source of semantic data. They enumerate some weaknesses of works that originated in paper form, the major being internal inconsistency in dealing with semantic phenomena. One could expect that in scholarly works which, as the *LMILP*, has been launched 60 years ago it should be even worse. Changing team composition, fluid editorial rules, successive chief editor changes, all that does not contribute to lexicographic coherence.

3 It is revealing to know, for example, that for about 1500 occurrences of the label *phil.*, «philosophy», only 340 are also determined as *termini technici*. 

---

---
One can ask why there are three parallel means of encoding linguistic phenomena that have at first glance so much in common. What seems to be at work here is the fact that the concept of technical terminology that emerges from between the lines of the *LMILP* is not clearly defined. Is terminology being considered – to cite the classification of Lothar Hoffmann⁴ – a functional language, a language variation, a sub-language or group language? Should the terms be defined by reference to the medieval communities which used them? Or maybe to the texts in which they were employed? Is it its frequency in special corpora or maybe lexicographer’s intuition and his knowledge of the medieval world that are crucial in determining what a term is and what it is not? Without answering those questions, one should not expect that term encoding could be coherent and the overlapping or redundant classification eliminated.

In case of the above-mentioned triple term/subject field labelling, it seems that there exists in the *LMILP*, so to say, an informal «technicality» scale. Words that are perceived by lexicographers as obviously technical are labelled *t. t.*; those for which lexicographer is not sure of their technical nature, but which are strongly associated with some topic, are defined by means of bare domain label; lastly, those for which only general subject of texts in which they occur can be defined are equipped with *in textibus* note.

### T. t. label use

The *terminus technicus* label was introduced in the first volume of the *LMILP*, but the dynamics of its use have changed through the years (see Table 1⁵).

<table>
<thead>
<tr>
<th></th>
<th>Vol. 1 (A-B)</th>
<th>Vol. 2 (C)</th>
<th>Vol. 3 (D-E)</th>
<th>Vol. 4 (F-H)</th>
<th>Vol. 5 (I-L)</th>
<th>Vol. 6 (M-O)</th>
<th>Vol. 7 (P-Q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of occurrences of t.t. label</td>
<td>268</td>
<td>328</td>
<td>295</td>
<td>123</td>
<td>267</td>
<td>139</td>
<td>201</td>
</tr>
<tr>
<td>(*t.t. label occurrences/estimated entries number) × 10000</td>
<td>506</td>
<td>563</td>
<td>645</td>
<td>360</td>
<td>491</td>
<td>370</td>
<td>323</td>
</tr>
<tr>
<td>number of subject field labels</td>
<td>10</td>
<td>12</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>10</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 1. Dynamics of the *t. t.* label use.


⁵ All the numbers given are only estimates, so the accurate calculation will be available once the proofreading and structural encoding starts.
The most extensive use of the label in question can be observed in vol. 2 and vol. 3 with approximately 300 occurrences, making about 600 occurrences for 10000 articles. As one can easily see, the use of the t. t. label was steadily growing within the first three volumes. At the opposite pole we find volumes 4, 6 and 7 with half as many occurrences. This downward trend, however, was reversed for a while in vol. 5 where 236 occurrences of t. t. can be found. This momentary change can be easily explained since volume I-L comprises a plethora of compounds beginning with prefix in- which one could expect frequent in coining of technical terms.

As has been mentioned, the terminus technicus label has been used solely with subject field indicators. Their number varies between 10 in vol. 1 and 6, and 15 in the volume 7. Here, the tendency seems to be opposite to the one discussed above, with number of labels growing except for volume 6.

Of the labels that are missing from the first volume, only log., standing for vox logica, and mus., vox musica, will be in common use in forthcoming volumes. Other labels absent from the first volume were introduced only on ad hoc basis, so they should be considered as, so to say, licentia lexicographica. This can be said, for instance, of opt. for vox optica which has two occurrences and num., numismata, with only one occurrence. On the other side, the label liturg. (four occurrences), which should have marked liturgical terms, while present in the abbreviations list which had preceded the first fascicle, has probably been subsumed by more general label eccl. The same is true for the label math. (nine occurrences) which may have lost its significance in favour of more specific (and useful) geom. (45 occurrences) or astr., since it is first and foremost optical and astronomical texts that supply mathematical content in the LMILP.

In what concerns label usage proportions the following remarks can be made:

1. Just two labels, iur. and phil., account for more than a half (57%) of all occurrences. If one agrees to consider log., «logic», a sub-field of philosophy, the impression of disproportion becomes even stronger, since now only three labels account for almost 70% of label use. This imbalance seems to be a reflection of the simple fact, that the major part of the LMILP’s sources are legal texts, namely privileges, judicial records and so on. Philosophical texts, on the other hand, form the largest part of academic sources. Another 10% is made up of occurrences of the label eccl. (= ecclesiastical technical terms), and 6% of grammatical words. So one can say that 85% of the occurrences of subject field labels (at least those accompanied by t. t.) comes from five of the 15 labels.

A quick glance at standardized data reveals other varieties of label use. One of the labels that is subjected to a most remarkable variation is med., «medicine», which, with only 13 occurrences in the first six volumes, experiences a rapid growth in use in the last volume with 17 occurrences (that is 60% of its

---

\(^6\) Naturally, the influence of the special interest that lexicographers and excerptors could have for these very domains cannot be excluded from the list of possible explanations.
total use). This is the volume where one of the most important Polish medical treatises, namely De practica by Thomas of Wrocław\(^7\) was introduced to the LMILP’s sources. The same is true for the subject field musica of which first occurrences can be traced back to the third volume, in which new musical sources were introduced for the first time\(^8\). The next major change can be noticed with regard to the label log., vox logica, of which the most extensive use can be observed in the third volume. The reason for that becomes clear when one takes into account the fact that the volume comprises such word families as definire, disiungere, distribuere, exponere, extremus.

Now I will proceed to more close examination of the way in which words labelled as technical are encoded in the LMILP.

**Technical terms in the macro- and microstructure of the LMILP**

The goal of the following analysis is twofold. Firstly, it aims to identify potential difficulties in extracting terminological knowledge from machine-readable dictionary, knowledge which is conveyed directly and indirectly by dictionary content and its structure as well. The second goal of the following is a metalexicographic one, since it will be shown that digitization projects provide lexicographers with an excellent opportunity to rethink their professional practice and critically examine the information with which they provide their users.

Two domain labels were chosen for analysis: geom., «geometry», and gramm., «grammar». From the total number of 150 entries, 49 are labelled as technical terms of geometry and 101 as grammar terms. There are many reasons why those precise subject fields had been picked. The first is a practical one: both have relatively small number of occurrences, which makes analysis of every occurrence feasible. Secondly, their distribution is of relatively small dispersion, which means that their use has not changed remarkably through time. Lastly, they represent relatively disparate subjects, which reduces the potential impact of, so to say, «same-lexicographer-effect»\(^9\). I will proceed through the macro- and micro-structural level of the dictionary.

---

\(^7\) Ed. T. J. Antry, Warszawa, 1989.


\(^9\) It is a well-known fact that in lexicographic teams there are always researchers who are specialists in one or many specific subject fields.
Technical terms in dictionary macrostructure

In the sample under analysis, there were 69 (45 %) entries for which the technical use or uses were the only one. The remainder (81, i.e. 55 %) consists of polysemous words with one or more senses labelled as technical. In this group, 18 entries (22%) have a technical sense coming first, which means (or at least should mean) its markedness in terms of semantic, chronological, or frequency priority. A closer look, however, reveals some intricate problems of term treatment in the LMILP.

In the entry *accentus*, for instance, the most evident (at least for classicists) meaning *distinctus modus pronuntiandi unam syllabam vocis* ranks as first on the list of senses. The remaining three meanings of the word have been grouped under one heading:

*accentus* … I. t. t. gramm. *distinctus modus pronuntiandi*. II. 1. … 2. … 3. …

Since the rules of sense ordering were not explained by the first editors in the dictionary introduction, information that underlies sense arrangement can sometimes be difficult to decipher. There is thus no way to know if the technical sense is first only in terms of its chronology or also in terms of frequency in medieval texts, also the semantic links between special and general meaning remain unclear. Did the general meaning evolve from the technical one according to the lexicographer? What is the shared conceptual area that enabled this semantic change? Whatever the precise answer is, the linear arrangement and grouping of senses force readers to conceive of frontal senses not only as serving as a basis for semantic alternations, but also as in a sense prototypical for a language user. In the LMILP this can lead to some doubtful, if not completely false conclusions.

For the language user described in the LMILP, for example, the geometrical meaning of the word *angulus, figura, quae ex duabus lineis in unum punctum desinentibus constat, « angle», is more evident than its material sense of *locus, ubi duo latera conveniunt atque in acutum desinunt, « corner»*. The same can be said about the entry *circulus*, in which the first sense is labelled as a technical term of geometry, or *derivatio* (1. *gram. t. t. vocum origo vel explicatio*. 2. *origo (sensu latiore), initium*) where the technical use seems to motivate the more general one.

Naturally, there are some exceptional cases in which the dictionary user will not be provoked to make any similar assumptions. One such border case is the entry *ellipsis*. Stemming from Greek ἔλλειψις, the noun has inherited its orig-

---

10 On the other hand, it is a standard entry arrangement of the lexems for which the non-technical meaning is ephemeral, like in *barbarismus* … I. *gramm. t. t. sermonis vitium vocibus peregrinis contaminati*. II. *mores barbari*. 
Digital dictionary digitization as a lexicographic challenge

In the above example it is hardly probable that a less advanced dictionary user could look for traces of semantic motivation between two terms. Moreover, even very advanced dictionary users will have a hard time trying to invent a shared meaning that would explain inclusion of both terms under the same headword\(^\text{12}\), the sole argument for joining two \textit{de facto} homonyms in one entry being their common etymology\(^\text{13}\).

The chances for automatic sense retrieval may become even worse when the lexicographer decides not to draw a sharp distinction between senses by putting technical subsense(s) in the «notes» section of the main sense. This part of the entry, labelled as \textit{praec. (}=\textit{praecipue}, «in particular», collects bizarreries of use and form which otherwise would not be easily classifiable; it can also play the role of an \textit{ad hoc} container for shades of meaning or for senses that seem to the lexicographer purely textual and ephemeral, therefore not autonomous enough to be separated. We see such a structure in the entry \textit{defectivus}, for instance, where the technical meaning, \textit{in declinatione definiens}, is characterized as \textit{praecipue} of general \textit{debilis, imperfectus, vitiosus etc.} In this example the subordination of the technical term probably expresses its genetic filiation and semantic proximity to the main sense. What is more, senses labelled as particular can sometimes pass without any definition, inheriting that of the main sense. As an example, the entry \textit{infinitus} can be mentioned, in which technical, namely grammatical and logical, senses have been subsumed under the more general one (\textit{non definitus, incertus, communis}), without being defined separately:

\textit{infinitus ... non definitus, incertus, communis. Praec. a. gram. t.t. ... β. log. t.t. ...}

At first glance, it is not clear how we should interpret this lack of definitions. Is it a matter of typographic convention, a shorthand for more verbose and largely redundant «1. non definitus... 2. gram. t. t. non definitus... 3. log. t. t. non definitus»? Or does such an arrangement perhaps rest on an assumption that the lexical meaning of technical terms is no different from that of the general language word, since both differ only in their distribution, the subject

\(^{11}\) In Polish there is a single word, \textit{elipsa}, which serves both meanings.

\(^{12}\) Another and in my opinion better solution was adopted by \textit{Trésor de la langue française informatisé} [http://atilf.atilf.fr/] where two homonym entries were separated, i. e. 1. \textit{ellipse} and 2. \textit{ellipsa}.

\(^{13}\) An example of similar editorial practice is the entry \textit{circumflecto}, in which two different and semantically independent senses have been joined: \textit{circumflecto} ... I. gramm. t. t. \textit{accentu circumflexo ... notare (in Graecis vocibus)}. II. dep. genua circum aliquid flectere.
field of the sources, in which they are attested? Such an interpretation could be easily corroborated by other examples of similar nesting practices as can be found for instance in the entry *obliquo*:

*obliquo* … *obliquum (non rectum) reddere, flectere. Praec. geom. t. t. ….*\(^{14}\)

As for regular sense nesting, the user can gain some aid understanding the sense hierarchy when s/he has been given clues on the relations between technical and other senses. This function can be fulfilled by additional labelling of technical sense instances. Traditionally, one of the most widespread methods of sense arrangement is based on the opposition of literal and metaphorical uses of a word, which, in turn, largely relies on the assumption of genetic and to some extent also psychological precedence of literal over the figurative meaning. To a certain degree, its variation is another lexicographic totem, that is the opposition between concrete and abstract meaning, which in the *LMILP* is expressed in a pair of labels *abstr.* and *concr.*. Since their use is well established in traditional lexicography, they need also to be carefully taken into account when speaking of technical terms. It is true that this sort of opposition is easily retrievable from digital dictionaries, its usefulness in the description of terminology, though, remains limited, since that perspective is not applied in a systematic way either to technical vocabulary or to general language words. To get an idea of this kind of sense marking, one can refer to the example of *figura* the grammatical sense of which (*gram. t. t. de formis casuum, personarum, generum sim.*) was classified under *figurative use* (*transl.*), while its geometry meaning was subsumed under heading *propr.*. In what concerns the opposition concrete-abstract, an example of an adjective *contingens* can be recalled, in which the geometrical sense of *linea quae alteram (praec. circularem) contingit* is labelled as *concr.*, while the philosophical meaning « fortuitous » is marked as *abstr.*. In the small and thus unrepresentative sample which has been collected for the present study one can observe relative dominance of metaphorically derived senses over literal ones. Although it would not be reasonable to risk a general opinion, such a prevalence is of no surprise, when one takes account of the nature of both subject fields, i.e. *gram.* and *geom.* Moreover, one can expect similar results for the entire dictionary as the label set is rather science than craft oriented.

\(^{14}\) Similar arrangement can be found s. v. *oblique*, where a special meaning is introduced as a parenthetical note within general sense definition: *oblique* … *incline* … (*praec. geom. t. t. ap. Vitelonem*).
Technical terms in dictionary microstructure

One of the most easily retrievable pieces of information from a machine readable dictionary consists of data extracted from the beginning part of the entry. Being usually highly standardized and forming in an ideal case a set of fields (which could be called PoS field etc.), it can provide us with a treasure of data that may in turn be exploited as working hypotheses for further linguistic research.

The etymology and «other dictionaries» sections of the entries reveal that in the group of grammatical terms that have been collected about 70% are already attested in classical Latin, another 10% come from late Latin and the remaining 20% are a product of medieval word formation. As to the latter, it can be concluded that at least in the group in question the most productive pattern of word formation was coining new adverbs from classical adjectives (five occurrences). Examples of knowledge retrieval of this type can be multiplied. For instance, it is easy to discover that a half of geometry vocabulary is formed by substantives, 30% by adjectives, 12% by verbs. This proves that digital dictionaries can provide researchers with at least basic description of medieval Latin terminology and general vocabulary system as well.

As to term definitions, their form is anything but uniform in the LMILP, which can be a major obstacle when one endeavours to automatically retrieve sense relations between terms such as synonymy or hyponymy. The definitions represent almost all the possible formal types, so one can find there examples of analytical, synthetic, morphosemantic, and encyclopaedic definitions as well. The most widespread of these, however, is the analytical one. Nevertheless, it is hardly surprising that technical language definitions tend to be permeated by encyclopaedic information, a phenomenon which is in fact in accord with contemporary research. As Maria Teresa Cabré Castellví puts it, «cognitive psychology and philosophy have stressed the difficulty of drawing a clear separation between general and specialised knowledge and have shown how general knowledge contributes to the acquisition of specialised knowledge.» Moreover, it is in relation to technical terms that the traditional questions of lexicographical debate become even more challenging. How much encyclopaedia information then should be given to the dictionary user? What conceptual system should be reflected in lexicographic divisions, the medieval or the contemporary one?

As to the first question, the encyclopaedic character of term definitions is one of the problems that lexicographers have to cope with in their everyday practice.

15 This is a section in the LMILP in which one can find bibliographic abbreviations of the classical and medieval Latin dictionaries in which the word in question has its entry, e.g. Th. stands for Thesaurus Linguae Latinae and so on.
On the one hand, those whose background is in linguistics tend to feel themselves obliged to limit the amount of world knowledge, even if they consider it indispensable for understanding meaning. On the other, those with principal training in history, opt for a more in-depth account of the medieval world, although it can result in encyclopaedic rather than lexicographic scholarly work. As for definitions of actual technical terms in the *LMILP*, the problem should be addressed for Polish and Latin definitions separately.

In Polish definitions recourse is often made to an equivalent technical term which is then explained in general Polish to precise its meaning and, thus, help the user to grasp it. Therefore, if possible, the technical term definitions have been mostly given the following form:

*catachresis* ... *gram. t.t. Katachreza, figura polegająca na użyciu wyrazu w niewłaściwym znaczeniu* [= «catachresis, figure based on use of word in a wrong sense»]

Occasionally, lexicographers seem to realize that the Polish synonym is neither easily understandable nor widely used. This consciousness results in somewhat redundant, but thus even more significant use of the adjective *tak zwany*, 'so called', abbreviated as *tzw.* as in:

*conceptorius* ... *gramm. t. t. Oratio tzw. Sylepsa, konstrukcja gramatyczna...* [= «so called syllepsis, grammatical construction ... »].

The exactly opposite approach has been adopted by an author of the Latin definition of *evoco*:

*evoco* ... *gram. t. t. Aliquid efficere, alicuius rei causa esse*

Such sense explication, although in accordance with the postulates of general language lexicography, does not give a user any hint about why the sense in question should be even considered as technical. There is not the slightest suggestion of the medieval theory of *evocatio*, without which it seems improbable that the quotations that follow can be understood. Even the reference to the entry *evocatio*, although expected, is of no help, as that entry does not contain any further explication of the term either, *evocatio* being defined as *actus efficendi, evocandi aliquid*. It should be accentuated that this could be the correct definition of a general language word. Yet it is obvious that the author, although s/he was fully conscious of the technical meaning of the word, was compelled to take a purely linguistic stance and deprived his/her definition of any reference to the medieval world.

This inadequacy of typical lexicographic definition in term description has been at least partially tackled in the *LMILP* in a few ways:

a) One of them is the tactic of replacing the Latin definition of a term with that taken from the medieval source. It was employed for example in *obiectum*:
b) The other possible strategy is inclusion of antonyms and synonyms in the definition, as we see in *derivativus*:

*derivativus* ... *gram. t. t.* ... *qui ab altero derivatur (de partibus orationis, opp. primitivus).*

This is used first of all as a means of precising term meaning, but synonymy and antonymy relations seem sometimes to form such a crucial part of word meaning that their description actually precedes word's definition, example of which can be found in the entry *extrinsecus*:

*extrinsecus* ... *opp. intrinsecus I. de rebus ...*

or in the entry *explicite*:

*explicite* ... *phil. et gram. t. t. (opp. implicit) ... clare, plane, distincte, aperte, expresse.*

This strategy enables the user to place the term in the context of medieval knowledge structures and understand it as a part of the scientific discourse of that time.

Next to the definitions, it is the source quotation section which is most crucial for the determination of a word’s meaning. Although it would take too long to describe the *LMILP’s* quoting strategy in great detail, a brief analysis of one of its features can stimulate reflection on the problem.

First of all, it is worth noting that thanks to the source quotations it is possible to automatically trace the date of the first occurrence of the technical term. If in the list of sources the works are linked to geographical data, the spatio-temporal distribution of technical terms can also be easily extracted. There is, though, a remarkable limitation to this optimistic view, which is imposed by the *LMILP’s* editorial conventions. In the sample under analysis, 65 senses, that is 45% of the total, are not endowed with any source quotation, since they are not registered for word senses which were known in classical Latin. Therefore, it is not possible to put them precisely in chronological or geographical frames.

As to quotations which illustrate grammatical senses, the most widely represented source is Glog. Alex., a xvth century commentary on the *Doctrinale* of Alexander de Villa Dei written by Johannes Glogoviensis. The 50 remaining occurrences come mainly from works on grammar, rhetoric, logic, and theology. As for senses labelled as geometrical, they are attested in the greatest number in an optical treatise of Vitelo (21 quotations). The next two most cited sources,

---

17 Johannes Glogoviensis, *Prima pars doctrinalis Alexandri...*, Lipsiae, 1525; *Exercitium secunde partis Alexandri...*, Viennae, 1517.
18 *Vitellonis Thuringopoloni Opticæ libri decem...*, Basileae, 1575.
with only seven occurrences each, are Albert.\textsuperscript{19} and Martin.\textsuperscript{20}, the former being a commentary on an astronomical treatise, the latter a work on geometry. The remaining ten quotations derive from astronomic (Copern. which is our source only by means of a convention\textsuperscript{21}) and philosophical works.

Discrepancy between the domain to which the source belongs and the label attributed to the sense, although it may seem obvious regarding the interdisciplinary nature of the terms, sheds also light on the problematic nature of subject field classification of medieval terms and raises some essential questions. What do or what should subject field labels account for? Do gram. and geom. labels represent medieval knowledge system divisions or rather reflect modern reparation of research domains? The problem was most recently raised by Robert Martin\textsuperscript{22}, who mentions that in the Dictionnaire du Moyen Français only those subject fields were adopted which were already known in the Middle Ages. Although the LMILP’s labelling system is far from being coherent, there are some signs that the authors were often concious that modern divisions were not always the best ones. This awareness can be now traced in the terse commentaries that enhance the notion of the subject field label by referring to the medieval system of knowledge:

\textit{casualis ... gram. t.t. ... N. in grammatica speculativa.}

The other way to tackle the complex nature of medieval terminological fields is the use of several labels for one sense. In the sample under analysis, the number of senses to which two or more labels were attributed at once (e. g. log. et gram., geom. et astr.) is relatively high, with 18 occurrences. It appears, then, a common way to signalize the interdisciplinarity of medieval technical vocabularies.

\begin{flushright}
\textsuperscript{19} Albertus de Brudzewo, \textit{Commentariolum super theoricas novas planetarum Georgii Purbachii...} (ed. L. A. Birkenmajer, Kraków, 1900).
\textsuperscript{20} Martinus Rex de Premislia, \textit{Geometriae practicae seu artis mensurationum tractatus} (ed. L. A. Birkenmajer, Warszawa, 1895).
\end{flushright}
Conclusion

Machine readable dictionaries can be a valuable source of knowledge about linguistic phenomena, including terminology. Available methods of automatic or semi-automatic data mining provide us more rapidly than before with a plethora of structured information. In the not so distant future, we will certainly be able to generate a basic, schematic account of medieval vocabulary, taking machine readable dictionaries as a starting point. Certainly, this account will be imperfect and will require manual processing, nevertheless it will supply us with a number of research hypotheses on medieval Latin.

Yet, automatic data extraction should certainly be preceded by posing a question of whether machine readable dictionary is a source of dispersed linguistic observations or whether maybe the only insight it can give us is one into historically changing dictionary making practices.

Krzysztof Nowak
Lexicon Mediae et Infimae Latinitatis Polonorum
Institute of Polish Language (Polish Academy of Sciences)
krzysztofn@ijp-pan.krakow.pl

Abstract. — This paper presents the brief report of the feasibility of automatic term extraction from the Lexicon Mediae et Infimae Latinitatis Polonorum. Firstly, the representation of technical terms in the printed version of the LMILP is scrutinized. Next, the analysis of the preconceptions about the medieval terminological system inherent in the LMILP is provided. In the conclusions, the main problems of automatic term extraction from print dictionaries are demonstrated.

Résumé. — L’article présente les possibilités d’extraire automatiquement la terminologie technique du Lexicon Mediae et Infimae Latinitatis Polonorum. On examine d’abord la façon dont les termes techniques sont représentés dans l’ouvrage imprimé. On analyse ensuite les principes qui semblent fonder le traitement des termes techniques dans le dictionnaire. Enfin on montre les plus grands problèmes posés par l’extraction automatique de la terminologie.