



Aleksei Kelli

*Docent
of Intellectual Property Law
University of Tartu*



Arvi Tavast

*Ph.D.
Head of Language Technology
Institute of the
Estonian Language*



Heiki Pisuke^{*1}

*Head of Unit, General Secretariat
of the Council of the
European Union, Brussels
Professor of Intellectual Property Law
University of Tartu*

Copyright and Constitutional Aspects of Digital Language Resources:

The Estonian Approach^{*2}

1. Introduction

Language is the fundamental basis of national identity, self-determination, and culture. Linguistic diversity is a major guarantee of the cultural diversity of the world. Therefore, it is hard to overestimate the importance of the preservation and development of national languages. Language is also a tool of communication between people both within a state and internationally. Language constitutes an interdisciplinary domain of which legal issues form an important part, and language is a precondition for formulation and enjoyment of fundamental rights.

This approach has been recognised at the highest regulatory level in Estonia. Namely, the preamble of the Constitution of the Republic of Estonia^{*3} provides that: '[w]ith unwavering faith and a steadfast will to strengthen and develop the state [...] which shall guarantee the preservation of the Estonian nation, language and culture through the ages'. According to legal commentators, 'the inclusion of the protection of the Estonian language as a basic principle in the preamble of the Constitution implies its recognition as the core value of the nation. It is impossible to separate the Estonian culture from language'.^{*4}

The Estonian Constitution establishes an additional principle related to language. Pursuant to § 6 of the Constitution, '[t]he official language of Estonia is Estonian'.^{*5} It is explained that '[e]very society needs to

¹ The views expressed in this article are those of the authors and do not reflect the official opinion of the Council of the European Union.

² This publication has been supported by the European Social Foundation through the Research and Innovation Policy Monitoring Programme.

³ Eesti Vabariigi põhiseadus (as entering into force on 3.7.1992). – RT 1992, 26, 349; RT I, 27.4.2011, 1 (in Estonian). English translation available via <http://www.legaltext.ee/> (most recently accessed on 11.3.2012).

⁴ R. Narits, H. Schneider. Preambul (Preamble). – Eesti Vabariigi põhiseadus. Kommenteeritud väljaanne. Teine, täiendatud väljaanne (Constitution of the Republic of Estonia, Commented Edition, Second Edition), edited by a panel led by E.-J. Truuväli. Tallinn: Juura, Õigusteabe AS 2008, p. 35 (in Estonian).

⁵ This Constitutional provision is implemented by the Language Act, which is intended to 'develop, preserve and protect the Estonian language and ensure the use of the Estonian language as the main language for communication in all spheres of

have normal communication and social cohesion among its members. This, however, can be achieved only if there is a shared language. According to the Constitution, the shared language in the Estonian society is Estonian'.⁶

New information and communication technologies (ICT), as well as the entirety of progress of technology, can influence the development of language. This can be illustrated by the fact that, in increasing numbers, conversation partners who use the Estonian language are artificial technical systems. Ranging from user interfaces of mobile phones to central information services, these systems also include safety-critical applications such as medical devices and vehicles' driver information systems. None of these systems at their present level of development really understands or speaks natural languages in the human sense. Instead, they use sophisticated technology to emulate human language behaviour and extract unambiguous information from human input. All of these language technologies are built upon digital language resources.

For the purpose of this article, digital language resources are defined as databases whose content consists of many written and oral texts. Selection of the texts to be included in this database is a creative process that requires collection and systematisation of the material for inclusion in this database. Therefore, the database is protected by copyright as a work⁷ under the Estonian Copyright Act.⁸ In this article, 'digital language resources' refers to the content of the database, a collection of written and spoken texts, and not to technical tools and Web applications.

The challenges related to creation of new language technologies are not limited to technological problems. There is also a myriad of legal issues, ranging from personal data protection to intellectual property problems, first and foremost in the field of copyright.

The aim of this article is to study some issues related to the development of digital language resources, which constitute a crucial challenge for new language technologies. For establishment of digital language resources, there is a need to utilise texts in enormous quantities and great variety. Most of these texts are protected by copyright as literary works.⁹ According to copyright rules, the use of works can either be based on right-holders' consent in the form of a contract or follow the rules of free use of works or, in other words, copyright exceptions to the exclusive rights of the author established by the legislator. The first option is called the licensing model and the second one the exception model. The practice of different countries in relying on the licensing or exception model to develop digital language resources varies, depending on policy and legal considerations and on the size and structure of the local market.

The creation of national digital language resources in Estonia serves public interests. Such a digital language resources database is created not for direct commercial purposes. The aim is to fulfil the Constitutional task of preserving and developing the Estonian language. Therefore, in this article, the authors concentrate on the issues of applicability of exceptions to exclusive rights of the author in the Estonian Copyright Act, which facilitates the creation of digital language resources.

The research related to copyright aspects of digital language resources is in its initial stage in Europe. This is a field of knowledge that requires an interdisciplinary approach and joint effort of linguists, lawyers, and IT specialists. The present article reflects the research results from the Estonian state programme for language technology and the research and innovation policy monitoring programme. The research in this interdisciplinary field is undertaken by scientists from the Institute of the Estonian Language and the Faculty of Law of the University of Tartu.

The topic of this article is directly tied in with several European Union policies. One of the aims of the European Union is to promote multilingualism. The overall goal is 'to contribute to a truly integrated, borderless digital Single Market by ensuring easy access to online services and creating better conditions for the development and use of rich content in Europe's many languages. The end result will be a digital economy, a society where knowledge and skills as well as online services, both public and private, can flow

public life' (§1). See Keeleseadus (as entering into force on 1.7.2011). – RT I, 18.3.2011, 1 (in Estonian). English translation available via <http://www.legaltext.ee/> (most recently accessed on 13.3.2012).

⁶ K. Merusk *et al.* §6. – Eesti Vabariigi põhiseadus. Kommenteeritud väljaanne (see Note 4), p. 72 (in Estonian).

⁷ Another option is to protect databases through sui generis rights provided by Directive 96/9/EC of the European Parliament and of the Council of 11.3.1996 on the legal protection of databases and Chapter VIII1 of the Estonian Copyright Act.

⁸ Autoriõiguse seadus (as entering into force on 12.12.1992). – RT 1992, 49, 615; RT I, 28.12.2011, 1 (in Estonian). Unofficial translation available via <http://www.legaltext.ee/> (most recently accessed on 11.3.2012).

⁹ To some extent, problems related to management of digital language resources are not entirely unique. For instance, mass digitalisation projects (e.g., by libraries) often face the same obstacles.

freely across national and language borders'.^{*10} Therefore, the Member States co-operate with the aim of sharing and trading language resources, both content data and digital tools.^{*11} In this article, the authors concentrate primarily on national issues of creation of digital language resources.

In the following sections, the authors first explain technological and copyright-related aspects of digital language resources. After that, they consider limitations of an author's exclusive rights—in other words, the free-use provisions of the Estonian Copyright Act, which could serve as a legal basis for development of national digital language resources. These copyright exceptions are analysed in light of the Constitutional guarantee of preservation of the Estonian language and in the conceptual framework of the 'three-step test', which is a fundamental concept for limitation of authors' rights.

2. Some technical and copyright issues of digital language resources

Language can be supported by technology on a variety of levels. Accented characters, sorting orders, and date formats are already taken for granted. Language-technology applications that are available for the Estonian language include spelling checkers, machine translation, speech synthesis (converting written text into speech), and speech recognition (converting speech to written text). There is a noticeable difference in quality between these tools and similar tools for major languages. Given the market sizes, it is natural for English to be the first language in which the cutting edge of technology is developed. For instance, conversation agents such as Apple's Siri and intelligent information retrieval were 'born' in English. An example of the latter is the IBM Watson computer system^{*12}, which in 2011 defeated human champions in an episode of the quiz show *Jeopardy*^{*13}, which is based on contestants answering questions posed in natural language.

While it would be theoretically possible to program such systems explicitly with language competence, this task would be prohibitively time-consuming and error-prone in practice. Also, the entire task of coding language rules would have to be repeated for each new language, making the effort non-scalable and especially infeasible for languages with smaller numbers of speakers, which, in turn, makes fewer resources available for language-technology development.

Modern statistical methods employ a different approach by shifting the emphasis from program logic to language data.^{*14} The program is mainly concerned with implementing a machine learning algorithm, which then enables the system to learn natural languages more or less as a human being would have, only on a more formal level. To this end, the system needs to be provided with numerous authentic examples of how humans use the language in question. The quantity of such language data is correlated with the quality of the resulting language-technology system. The current worldwide trend is to keep increasing the sizes of language resources, from hundreds of millions to billions of words. A billion words corresponds to roughly 15,000 midsize printed books.

It is possible to distinguish between two types of machine learning: assisted and unassisted. In language technology, the difference lies in the kind of texts presented to the system. Unassisted learning uses normal texts as they are found: books, Web sites, chat transcripts, recordings of talk shows, and any others that can be made available to the machine in huge quantities. For straightforward topics such as learning of translational equivalents at sentence or phrase level, on the basis of existing human translations, this works fairly well, as confirmed by the constantly growing usefulness of machine translation systems such as Google Translate. However, even slightly more complex inferences require more sophisticated learning techniques and greater processing power than are currently available for mainstream language-technology

¹⁰ An inventory of Community actions in the field of multilingualism – 2011 update. Commission Staff Working Paper. Brussels, 7.7.2011, SEC(2011) 926 final, p. 26.

¹¹ For instance, Estonia is one of the founding members of the Common Language Resources and Technology Infrastructure (CLARIN).

¹² D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A.A. Kalyanpur, A. Lally, J.W. Murdock, E. Nyberg, J. Prager, N. Schlaefer, C. Welty. Building Watson: An overview of the DeepQA Project. – AI Magazine 2010/Fall. Available at <http://www.aaai.org/Magazine/Watson/watson.php> (most recently accessed on 5.4.2012).

¹³ Watson—a system designed for answers. An IBM White Paper. <ftp://public.dhe.ibm.com/common/ssi/ecm/en/pow03061usen/POW03061USEN.PDF> (most recently accessed on 5.4.2012).

¹⁴ D. Jurafsky, J.H. Martin. Speech and Language Processing. Prentice Hall 2009.

applications. For example, the relatively simple knowledge that in English ‘is’ and ‘were’ are both forms of ‘to be’ is very hard to extract from running text, unless human beings give the machine some explicit clue about it first.

Assisted machine learning, which is, in essence, learning by example, is more realistic for most types of linguistic knowledge at our present level of technological development. Human linguists annotate texts with the kind of tags that they want the machine to be able to apply, then feed these tagged texts into the system (this is called training), and the system quickly learns to perform the same job independently.

The texts to be included in language resources need to be from the areas of language use that the resulting system will be used for. If training is limited to texts with no copyright protection—i.e., legislative texts and fiction from previous centuries^{*15}—the system will not cope with the modern discourse of user manuals or medical-information systems.

Samples of both written and spoken language use are needed for development of language-technology systems. Spoken-language resources contain audio or multimedia recordings of speech of various types, including spontaneous conversations. Written-language resources include anything from legislative texts to chat-room transcripts. For most types of resources, the important parameters of the source material, in addition to its size, are the topic and register. The meaning or message of the text is not used by language-technology applications at their present stage of development. In the compilation of a language resource, the texts are just piled together into a huge database with specific search capabilities, possibly with added tagging. Although it may be technically possible to retrieve substantial portions of the source texts from language resources, this is not what they are designed and normally used for. They are intended only for extraction of information about how language works, by both human researchers and machine learning algorithms in language-technology applications.

From the copyright perspective, the creation of digital language resources involves the use of written and oral texts that are usually protected by copyright. Subsection 4 (2) of the Copyright Act defines protectable work as ‘any original results in the literary, artistic or scientific domain which are expressed in an objective form and can be perceived and reproduced in this form either directly or by means of technical devices’.^{*16} Subsection 4 (6) of the Copyright Act provides that ‘[t]he protection of a work by copyright is presumed [...]. The burden of proof lies on the person who contests the protection of a work by copyright’. In practical terms, this means that any text written or spoken by a human being^{*17} enjoys copyright protection and its use either requires authorisation or shall be based on limitations.

Creation of digital language resources includes collection and reproduction of copyright-protected and non-protected written and oral texts. There are two clear legal possibilities for creation of digital language resource databases from non-protected texts:

- 1) It is allowable to use works that are not protected, on account of the expiry of the term of protection (which is, as a rule, 70 years *post mortem auctoris*)^{*18}, and
- 2) It is possible under §5 of the Estonian Copyright Act, which provides that legislation, administrative documents, court decisions, and official translations thereof are not copyright-protected.

It should be mentioned that even after a work leaves copyright, authorship has to be honoured. This requirement is provided by Article 6*bis* of the Berne Convention and also integrated into the Estonian Copyright Act. It is necessary to distinguish between two distinct legal concepts here: authorship^{*19} and the right to authorship.^{*20} The only obligation of the user after the end of the term of copyright protection is to honour authorship by making reference to the author of the text used. Pursuant to §44 (1) of the Copyright Act,

¹⁵ These two account for a large part of the material in current Estonian language resources; see <http://www.cl.ut.ee/korpused/> (most recently accessed on 5.4.2012).

¹⁶ The definition is based on Article 2 (1) of the Berne Convention for the Protection of Literary and Artistic Works. Berne, 9.9.1886 (which entered into force with respect to Estonia on 26.10.1994). – RT II 1994, 16–17, 49 (the Berne Convention).

¹⁷ One of the underlying concepts of Estonian copyright law reads: ‘The author of a work is the natural person or persons who created the work’ (§28 (2) of Copyright Act). In view of this concept, it is hard to say whether a work generated by a computer program is protectable or not. One approach would be to vest copyright in the person who created the program.

¹⁸ The copyright terms are unified within the EU, and, therefore, all EU countries have a similar situation. See Council Directive 93/98/EEC of 29.10.1993 harmonizing the term of protection of copyright and certain related rights. – OJ L 290, 24.11.1993, pp. 9–13.

¹⁹ Authorship is a connection between the work and its creator, and it is protected forever. See §44 of Copyright Act.

²⁰ The right of authorship involves a general term of protection, as a rule, of the life of the author plus 70 years. See §12 (1) 1) of Copyright Act.

‘[t]he authorship of a certain work, the name of the author and the honour and reputation of the author shall be protected without a term’^{*21}. As a matter of fact, the right to authorship is also enshrined in §39 of the Estonian Constitution^{*22}, which provides that ‘[a]n author has the inalienable right to his or her work. The state shall protect the rights of the author’. Compliance with this requirement in the context of development and management of digital language resources is not complicated or burdensome.

The creation of language resources cannot rely solely on materials outside the scope of copyright protection. In the process of development and preservation of language, it is crucial to use contemporary texts as ‘raw material’. These texts include works and extracts thereof that are still under copyright protection. The utilisation of copyright-protected material, however, can be based either on right-holders’ direct consent or on limitations of right-holders’ exclusive rights. Such limitations are foreseen in Chapter IV of the Estonian Copyright Act, ‘Limitations on Exercise of Economic Rights of Authors (Free Use of Works)’. In this article, we refer to these limitations also as copyright exceptions.

3. The application of copyright limitations for development of digital language resources

Copyright laws are not drafted in consideration of rapid technological developments.^{*23} Therefore, it has been established practice throughout the history of copyright law to try to interpret existing provisions in light of new developments, before corresponding changes to the law have been introduced. Now is not the first time when technological progress has been in tension with intellectual property (IP) systems. Often, IP experts have emphasised that technological advances have a destabilising effect on IP systems.^{*24} From our point of view, it is not so much IP systems that are adversely affected by technological advances. *Vice versa*, the IP system is not a thing in itself. Its main purpose is to enhance innovation in all spheres of life, including culture and social welfare—and also language. As far as the utilisation of the emerging IT technologies is concerned, the IP system more often constitutes a barrier to them than a mechanism to leverage unforeseen opportunities offered by these technologies.

This, however, does not mean that language resources cannot be developed within the current copyright regulations. It is essential to bear in mind that different exploitation methods determine whether one should utilise the exception or licensing model. In most cases, digital language resources can be developed and utilised in accordance with the existing copyright limitations (i.e., under the exception model). All copyright limitations, including those relevant to development of digital language resources, have to be placed within the context of the three-step test, which is a key in how to conceptualise, interpret, and implement the existing framework of copyright limitations in the Copyright Act.

The three-step test has its origins in the Berne Convention and is expressed there as follows: ‘[i]t shall be a matter for legislation in the countries of the Union to permit the reproduction of such works in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author.’^{*25} This provision entitles states signing the Berne Convention to limit the reproduction right under certain conditions. Similar regulation was included in the Agreement on Trade-Related Aspects of Intellectual Property Rights^{*26} (the TRIPS

²¹ According to §12 (1) 1) of Copyright Act, the right to authorship means to ‘appear in public as the creator of the work and claim recognition of the fact of creation of the work by way of relating the authorship of the work to the author’s person and name upon any use of the work’.

²² The Estonian Constitution has several provisions that address intellectual property protection. Section 39 protects creators’ (authors and inventors’) personal rights, and §32 covers, in addition to tangible property, economic rights and interests of IP-owners.

²³ On the EU level, there is Directive 2001/29/EC of the European Parliament and of the Council of 22.5.2001 on the harmonisation of certain aspects of copyright and related rights in the information society. – OJ L 167, 22.6.2001, pp. 10–19 (the InfoSoc Directive). This directive, however, does not introduce a new copyright paradigm. It instead specifies and complements the existing copyright regulations.

²⁴ See W. Cornish, D. Llewelyn. *Intellectual Property: Patents, Copyright, Trade Marks and Allied Rights*, Sixth Edition. London, Sweet & Maxwell 2007, p. 34.

²⁵ Article 9 (2) of the Berne Convention.

²⁶ Agreement on Trade-related Aspects of Intellectual Property Rights. Annex 1C of the Marrakesh Agreement Establishing the World Trade Organization. Marrakech, 15.4.1994. This entered into force with respect to Estonia on 13.12.1999 (cited in RT II 1999, 22, 123).

Agreement), which is also applicable in respect of Estonia. Pursuant to Article 13 of the TRIPS Agreement, 'Members shall confine limitations or exceptions to exclusive rights to certain special cases which do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the right holder'. The meaning of the provision is explained in the literature as follows: '[A]ny limitations imposed nationally to any exclusive rights granted under TRIPS, must satisfy the three-step criteria. Thus, the three-step test has been transformed into a general litmus test for domestic limitations on copyrighted works. The three-step test is not a public interest limitation to exclusive rights. Instead, it is a limitation on the *scope of limitations* that member states can implement to promote access and dissemination of works domestically. In sum, what appears to be a limitation to copyright, is actually [a] limit on the discretion and means by which member states can constrain the exercise of exclusive rights.'²⁷

There are certain differences in the conceptualisation of the three-step test between the Berne Convention and the TRIPS Agreement. According to commentaries on the TRIPS Agreement, 'while the Berne Convention only refers to the "reproduction" right of literary and artistic works, Article 13 of the TRIPS Agreement applies to all exclusive rights conferred'.²⁸ As a result, the TRIPS Agreement further limits the freedom of WTO countries to introduce new copyright exceptions to address challenges brought by technological change.

The Estonian Copyright Act has incorporated the concept of the three-step test²⁹ in the following wording: '[P]rovided that this does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author, it is permitted to use a work without the authorisation of its author and without payment of remuneration only in the cases directly prescribed in §§ 18–25 of this Act'.³⁰ It is necessary to explain that, upon inclusion of the three-step test in national copyright acts, its function is transformed. In the Berne Convention and the TRIPS Agreement, it is meant to limit countries' freedom to introduce copyright limitations in their national legislation. In copyright legislation such as the Estonian Copyright Act, its role is to guarantee that the author's rights are not violated even in cases in which use of a copyright-protected work is formally covered by an exception, where it still has an extremely adverse impact on the author's legitimate interests and there are no justifying circumstances. Below, the authors analyse the interaction of specific elements of the three-step test with the development of digital language resources. It is necessary to remember that the three conditions in the three-step test are cumulative.³¹

Next, the authors address copyright exceptions applicable in the process of creation of digital language resources. Since it is not enough to analyse specific exceptions if one is to determine their appropriateness, the authors place them within the framework of the three-step test. Therefore, the authors start by outlining conditions of specific exceptions such as the quotation right and the research exception. After this, the paper explores whether utilisation of these exceptions is in conflict with the normal exploitation of the work. Finally, arguments are presented to support the use of the quotation right and research exception to develop language resources. Constitutional guarantees for preservation of the Estonian language serve here as a central justification of development of digital language resources.

The first requirement of the three-step test is that an exception be directly prescribed in the Copyright Act. There are several exceptions under which the use of copyright-protected works is allowed without the consent of the author and without payment of remuneration. Section 19 of the Estonian Copyright Act sets forth several exceptions allowing 'free use' for certain scientific, educational, informational, and judicial aims. The exceptions relevant to digital language resources are the right to quote³² and use of works for

²⁷ R. Okediji. *Fostering Access to Education, Research and Dissemination of Knowledge through Copyright*. UNCTAD-ICTSD Dialogue on Moving the Pro-development IP Agenda Forward: Preserving Public Goods in Health, Education and Learning (2004), pp. 3–4. Available at http://www.iprsonline.org/unctadictsd/bellagio/docs/Okideiji_Bellagio4.pdf (most recently accessed on 17.3.2012).

²⁸ C.M. Correa. *Trade Related Aspects of Intellectual Property Rights: A Commentary on the TRIPS Agreement*. Oxford: Oxford University Press 2007, p. 136.

²⁹ Article 5 of the InfoSoc Directive also includes the three-step-test regulation.

³⁰ Copyright Act, §17.

³¹ See S. Ricketson, J.C. Ginsburg. *International Copyright and Neighbouring Rights: The Berne Convention and Beyond*, Second Edition, Volume I. Oxford University Press 2006, p. 763.

³² On the international level, the right to quote is regulated by the Berne Convention, which provides that '[i]t shall be permissible to make quotations from a work which has already been lawfully made available to the public, provided that their making is compatible with fair practice, and their extent does not exceed that justified by the purpose' (Article 10 (1)).

scientific purposes.^{*33} Both exceptions offer opportunities for, and limitations to, development of digital language resources.

In order to exercise the quotation right, one need not seek the author's authorisation or pay him or her remuneration. However, it is required to indicate the source, use only works that have been lawfully published, convey the idea of the work correctly, and confine quotations to reasonable limits.^{*34} Fulfilment of the requirements to use lawfully published works, make reference to the source, and convey the author's ideas correctly is not complicated, even though in the Internet era it is sometimes hard to say whether a work has been lawfully published.

The most complex issue is that the extent of included quotations must not exceed that justified by the purpose.^{*35} According to legal commentators, no limitation is placed on the amount that may be quoted. The concept of 'quotation' usually suggests that the thing quoted is part of a greater whole rather than the whole itself. Still, quotation of the whole work may be justifiable.^{*36} Within the context of development of digital language resources, the extent of quotation depends on the type of text. If a work is relatively short, it may be included in its entirety in the database of digital language resources. It is very hard to state an approximate percentage. The authors believe that this criterion is evaluated within the second and third requirements of the three-step test.

Since copyright law allows commercialisation of new works that contain quotations from works of other people, digital language resources made up of quotations can be sold or licensed for a fee.

Use for scientific purposes (i.e., under the research exception) is not to involve commercial exploitation of language resources. It also has to be limited to the purpose of illustration or to the extent justified by the purpose at educational and research institutions. According to copyright experts, '[b]oth illustration for teaching and scientific research must be the sole purpose of the use for which the exclusive rights may be restricted. Accordingly, when the reproduction or other use also fulfils an additional purpose, the exception or limitation must not apply'.^{*37} Here again, the validity and relevance of these purposes are evaluated within the second and third requirements of the three-step test.

The second requirement of the three-step test is that the use not conflict with the normal exploitation of the work. According to the explanation by copyright experts, 'whether the exempted use would otherwise fall within the range of activities from which the copyright owner would usually expect to receive compensation [...] "normal exploitation" will therefore require consideration of potential, as well as current and actual, uses or modes of extracting value from a work'.^{*38}

In this context, it is useful to refer to a case brought before a WTO dispute resolution panel in relation to interpretation of three-step test provisions in the TRIPS Agreement (described in a WTO report). This case provided the main official legal analysis of the scope and conditions of the three-step test and offers further insights into this matter. According to the WTO report, 'an exception or limitation to an exclusive right in domestic legislation rises to the level of a conflict with a normal exploitation of the work (i.e., the copyright or rather the whole bundle of exclusive rights conferred by the ownership of the copyright), if uses, that in principle are covered by that right but exempted under the exception or limitation, enter into economic competition with the ways that right holders normally extract economic value from that right to the work (i.e., the copyright) and thereby deprive them of significant or tangible commercial gains'.^{*39}

The preliminary market analysis of exploitation-models practice by copyright-holders of written and oral texts in Estonia reveals that the value is extracted mostly through selling of the texts as literary works (books, journals, e-books, etc.) or offering of advertising on a Web site or blog that contains copyright-protected texts, or these texts are not commercialised at all. The authors are unaware that any business actor holding copyright to many written texts would commercially exploit them by means of development and utilisation of digital language resources. The WTO report also emphasises that 'the extent of exercise or non-exercise of exclusive rights by right holders at a given point in time is of great relevance for assessing

³³ Copyright Act, §19.

³⁴ Clause 19 1) of Copyright Act.

³⁵ Clause 19 1) of Copyright Act.

³⁶ Ricketson, Ginsburg (see Note 31), p. 788.

³⁷ M.M. Walter, S. von Lewinski. *European Copyright Law: A Commentary*. Oxford University Press 2010, p. 1044.

³⁸ Ricketson, Ginsburg (see Note 31), p. 769.

³⁹ WTO. Report of the Panel. WT/DS160/R (15.6.2000), p. 48. Available at http://www.wto.org/english/news_e/news00_e/1234da.pdf (most recently accessed on 17.3.2012).

what is the normal exploitation with respect to a particular exclusive right in a particular market⁴⁰. Therefore, the authors assert that creation of digital language resources for scientific purposes is not in conflict with the normal exploitation of the work.

The third requirement of the three-step test is that the act not unreasonably prejudice the legitimate interests of the author. According to copyright experts, '[t]he words "not unreasonably prejudice" therefore allow the making of exceptions that may cause prejudice of a significant or substantial kind to the author's legitimate interests, provided that (i) the exception otherwise satisfies the first and second condition [...], and (ii) it is proportionate or within the limits of reason, i.e., if it is not unreasonable'⁴¹. The WTO report further elaborates on this matter: '[P]rejudice to the legitimate interests of right holders reaches an unreasonable level if an exception or limitation causes or has the potential to cause an unreasonable loss of income to the copyright owner.'⁴² This criterion requires one to consider whether development of digital resources of the Estonian language takes precedence over right-holders' economic interests in restriction of the quotation right and the research exception. The question has to do with public interest, which is country-specific. Legal scholars have characterised it as follows: "'Public interest', however, is a shifting concept that requires a careful balancing of competing claims in each case and one that is frequently interpreted in different ways at the national level, depending upon historical, cultural, and social circumstances"⁴³.

In Estonia, our language is protected at the Constitutional level. This means that the Constitutional guarantee of preservation of the Estonian language is a key in construction and implementation of the Estonian Copyright Act and a basis and justification for development of digital language resources.

Summary of conditions for use of the exception model to develop digital language resources

Name of the exception	Quotation right Copyright Act, §19's clause 1	Research exception Copyright Act, §19's clauses 2 and 3
Requirements of the Copyright Act	1) no need for authorisation and payment of remuneration; 2) reference to the source; 3) quoting from a lawfully published work in an extent that does not exceed that justified by the purpose; 4) commercial exploitation as allowed; and 5) exercise of the quotation right for non-profit and commercial purposes	1) no need for authorisation and payment of remuneration; 2) reference to the source; 3) the use of a lawfully published work for the purpose of illustration for scientific research to the extent justified by the purpose; 4) reproduction from a published work for the purpose of teaching or scientific research to the extent justified by the purpose at educational and research institutions; and 5) commercial exploitation as not allowed
Potential conflict with normal exploitation of the work	Right-holders of literary works do not usually commercially exploit their works to create digital language resources in Estonia. Development of digital language resources is not in conflict with normal exploitation of the works.	
Unreasonable prejudicing of the legitimate interests of the author	An important consideration in determination of whether the creation of digital language resources unreasonably prejudices the right-holders' legitimate interests is the need to preserve the Estonian language.	

⁴⁰ *Ibid.*, p. 50.

⁴¹ Ricketson, Ginsburg (see Note 31), p. 776.

⁴² WTO (see Note 39), p. 59.

⁴³ Ricketson, Ginsburg (see Note 31), p. 756.

4. Conclusions

Language is a *conditio sine qua non* for every culture; i.e., without language there is no culture. Language is not a static phenomenon. In contrast, its character is rather dynamic and it is continuously evolving in response to trends in society, culture, and technology. Advances in information and communication technologies enable the enhancement of language technologies. These technologies, however, are based on digital language resources, whose creation requires the utilisation of numerous copyright-protected texts. Therefore, copyright is an important issue in the development of digital language resources.

The official language of Estonia is Estonian, and the Constitution and other legal acts set in place measures to enforce this Constitutional provision. Therefore, the creation of digital language resources databases fulfils certain important aims of public law. At the same time, the exclusive economic and personal rights of an author or other right-holder are protected at the Constitutional level and more specifically by private law, in the Copyright Act. It is not easy to find the right balance between public- and private-law measures and among the divergent interests of the various stakeholders.

There are two main methods for creation of digital language resources databases: on the basis of a contract with copyright-holders (licensing model) and reliance on the provisions for fair-use exceptions (limitations to exclusive rights of the authors) in the Copyright Act (exception model).

In this article, the authors have explored whether copyright exceptions allow the development of digital language resources without acquisition of authorisation from right-holders and payment of remuneration. These exceptions were placed within the conceptual framework of the three-step test and interpreted in light of the Constitutional guarantee to preserve the Estonian language.

The authors' main conclusion is that the Estonian Copyright Act as constructed and implemented in light of the Estonian Constitution allows creation of digital language resources by exercise of the quotation right and the research exception. This conclusion is supported by the fact that the development of language resources is not in conflict with the normal exploitation of works. That is, right-holders for these texts are not commercially exploiting their works to create language resources. These activities are conducted mostly by public research institutions. The reason is that, because of the small number of individuals speaking the Estonian language, it is not economically sustainable for profit-oriented entities to invest in the development of digital language resources. Therefore, research institutions developing and utilising digital language resources are not depriving the right-holders whose texts are included in the resources of any revenue and there are no adverse financial consequences for right-holders.

It is also crucial to consider the Constitutional guarantees made for the Estonian language. The three-step test that constitutes the main standard for determination of whether a right-holder's interests and rights are violated allows limiting of these rights in cases wherein there is compelling reason for so doing. The authors conclude that the need to develop and preserve the Estonian language qualifies as compelling reason.