

SOME OBSERVATIONS ON ARMENIAN ELECTRONIC PROOFREADING SYSTEMS

Sargsyan M. A. PhD

Armenia, Yerevan, NAS RA Language Institute

DOI: https://doi.org/10.31435/rsglobal_ws/12062018/5888

ARTICLE INFO

Received: 30 April 2018

Accepted: 25 May 2018

Published: 12 June 2018

KEYWORDS

electronic proofreading,
proofreading system, spelling
mistakes, grammatical errors,
spellchecker

ABSTRACT

The article is devoted to a comprehensive study of Armenian electronic proofreading systems. As a result of the analysis, it is shown that, although a number of electronic proofreading systems have been created recently, none of them performs full proofreading. The author of the article has selected a passage of a scientific text, based on which a sample sentence has been composed, in which intentional spelling mistakes have been made in order to verify the possibilities of proofreading systems and to identify the shortcomings. A sample sentence has been entered in all Armenian electronic proofreading systems and checked. It was entered in the following systems: HySpell, Spellchecker, Translit spellchecker, SpellStar, PowerSpell, Add-Ons spellchecker. Studying all kinds of resources, comparing them with one another, and also identifying existing problems and shortcomings, in the conclusion, the author suggests developing new principles of the proofreading system for more comprehensive and full proofreading.

Copyright: © 2018 Sargsyan M. A. This is an open-access article distributed under the terms of the **Creative Commons Attribution License (CC BY)**. The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Introduction. A number of electronic proofreading systems and spell checkers have been created recently and are still being created. Now, we can say that the Armenian spell checkers are available on the Internet. However, it should be noted that none of them performs full proofreading. Not only do these programs fail to serve their purpose, to check spelling and grammar mistakes, to correct even the most frequent mistakes in texts but also they sometimes lead to arbitrary errors. Therefore, we can surely say that the quality of the Armenian Spell Checkers is low. It refers to both the volume of the vocabulary included in those programs and grammar, as far as not all types of grammatical errors can be corrected by them. All Armenian spell checkers have been created only by programmers and not by linguists. This may be considered the first shortcoming of the mentioned programs as there is no guarantee that the corrections made by those programs are right.

Material and methods. We have selected a passage of a scientific text, based on which a sample sentence has been composed, in which intentional spelling mistakes have been made in order to verify the possibilities of proofreading systems and to identify the shortcomings [1]. A sample sentence has been entered in all Armenian electronic proofreading systems and checked. It was entered in the following systems: *HySpell*, *Spellchecker*, *Translit spellchecker*, *SpellStar*, *PowerSpell*, *Add-Ons spellchecker*.

Before proceeding with the investigation of these systems, let us introduce how all the electronic proofreading systems work. To apply, some of them need to be downloaded on the computer, added to MS Word, Excel, PowerPoint and used there, while others are available on separate websites. Unlike the previous ones, the last ones allow to correct any electronic text directly on the electronic platform. Besides, they are profitable for the creators of the system, as they give opportunity to continuously enrich and develop the system. There are also such proofreading systems which operate in two ways, such as the foreign-language program Grammarly [2]. Some of the Armenian electronic proofreading systems and spell checkers investigated by us have similar opportunities, but there are some issues related to the launching. We will refer to them below.

What refers to correcting spelling and grammatical errors, the proofreading systems are implemented as follows: checking is done by comparing with the existing keywords in the database. If the checked word does not exist in the database, sometimes the system offers to import it into the dictionary. If the checked word is considered to be incorrect by the system, several options are offered, according to a formal similarity [3-4]. The number of versions varies in different correction systems. In most cases, 5-6 options are offered.

Among the Armenian electronic proofreading systems *Armenian PowerSpell* and *HySpell* programs are especially widely used. These are also relatively better proofreaders.

The *HySpell* [5] proofreading database includes about 158,000 word- units from several dictionaries: "New Dictionary of the Armenian Language" ("Nor Bargirk Haykazean Lezui"), E. Aghayan's "Modern Armenian Explanatory Dictionary" and St. Malkhasyan's "Armenian Explanatory Dictionary". On the one hand this can be considered good, as far as both the words in Eastern and Western Armenian are included, while on the other hand, if we separate Western Armenian words, the number of Eastern Armenian words will be reduced. It is not difficult to understand that, if the database is small, the proofreading will not be complete. This is probably one of the major drawbacks of the system. Additionally, there are problems with downloading and running. It is possible to download for free only for the Fire Fox browser, but it should be noted that today the usage of this browser is inappropriate because it is almost not used. The system has also easy-to-use versions adapted to Word 2007/2010. They work with different operating systems, but they are chargeable, which is also unreasonable, especially when the same database with a limited vocabulary is used in case of these versions, not including contemporary and new words, borrowings, foreign words, and so on.

The proofreader *Armenian PowerSpell* [6] was created for MS Word, Excel, PowerPoint. Let us note that it was easily downloaded and used in the operating system Windows 7, but it was not possible to run in Windows 8.1. This is, naturally, a major error because it shows that this system cannot be used in new versions of operating systems which are now more frequently used. Besides, this system is chargeable and can be used free of charge only once within 7 days.

The authors of the program *Armenian PowerSpell* note that their database includes 1,400,000 words. Our studies showed that in counting the vocabulary of the database, morphological forms, rather than the headwords were taken into account. This means that actually, the vocabulary is limited. Although unlike other systems, this system is based on the maximum number of dictionaries, we must note that a number of new words and neologisms are not included in the database. Even the fields of barbarisms, abbreviations, acronyms and proper names are not perfect. Thus, about 1000 proper names are noted, whereas, based on the recent accounts, the geographical names taken only are about 200,000 [7].

The database of the given system consists of two parts: mother and additional dictionaries. The additional dictionary includes proper names and abbreviations, which are rather incomplete. The main part of the database is represented in the mother dictionary. Both the mother dictionary and the additional one can be freely edited by any user. It should be noted that from the professional point of view it is not acceptable because it can be a cause of arbitrariness and mistakes.

As the authors of the *Armenian PowerSpell* proofing system note in their guide, this system allows you to check and correct the spelling, grammar, and punctuation errors. After reading the guide, the system description and finding out numerous grammatical, spelling, and punctuation errors, perhaps a question arises: whether or not the authors of this guidebook could be able to create a complete proofreading system.

The authors do not provide information about the dictionaries and database vocabulary of *Spellchecker* [8] and *Translit spellchecker* [9] proofreading systems. The *Spellchecker* offers two versions of proofreading: basic and advanced. The latter is chargeable. When we checked our sample-sentence with the *Spellchecker* and *Translit spellchecker* systems, some words were misinterpreted by the system, which revealed the deficiencies in the database and in the systems. Thus, the database does not have complete paradigms of systems of declension and conjugation, proper names, modern terminology, abbreviations and acronyms, barbarisms. Instead of presenting the forms of declension, the system offers incomprehensible and unsubstantiated options, such as separating an ending from a word-root with a spatial or a dash.

The *SpellStar* [10] proofreading system is probably the most incomplete system we have studied. For example, we intentionally entered several incorrect forms which were not even considered suspicious in this system. In general, it should be noted that not only does this program fail to serve its purpose, to correct spelling mistakes, to represent wordforms, variants according to a formal similarity, but it also offers unfounded, wrong, non-used options instead of correct words.

Add-ons spellchecker [11] is based on a small database. The vocabulary is limited: it contains only about 150000 wordforms. From the point of view of an application, this system is not expedient. First, the database is very poor, so the possibility of checking and correcting errors is also small. Additionally, this system is only for the Firefox browser which is the least used browser nowadays.

Research results. Summarizing the major shortcomings of the existing spell-checkers, let us mention some of them:

– *The vocabulary* of the mentioned programs is extremely incomplete, limited and does not reflect the current state of the Armenian vocabulary.

1. The programs are based on dictionaries that include many archaisms, rarely used words, but there are no neologisms and barbarisms in the database. It should be noted that today the latter is frequently used.

2. The programs do not have proofreading options for proper names, as their database does not include proper names, abbreviations and acronyms.

3. Some proofreading systems include open source public proofreading fields, but this is a mistake from professional perspective because it can become a basis for arbitrary mistakes.

4. The professional vocabulary is mostly not included in the existing spell checkers.

5. While calculating the database vocabulary of some programs, the morphological forms were taken into account, rather than the headwords. This clearly shows that the vocabulary of those programs is actually limited. For example, if a noun with singular and plural numbers can have 10 morphological forms and a verb can have 80 positive and 80 negative forms, totally 160 conjugated forms, then it is not difficult to count that the vocabulary is sufficiently reduced.

– *Grammatical errors* are inaccurately corrected because those programs:

1) do not include all the forms of the systems of declensional and conjugational paradigms in the database,

2) suggest unreasonable, incorrectly-used variants, such as separating an ending from a word-root with a spatial or a dash. There are no such variants in the Armenian language, except from the professional texts where the endings can be separated.

– The available spell checkers have *limited usage* because:

1. To use the mentioned programs, you have to buy a full package of the software, add in Microsoft Word, Excel, PowerPoint text editors and use them in your PC. This implies that only a limited number of people can use them. Moreover, even the creators cannot later modify or replenish their own software and correct errors made in those programs.

2. Free download option is offered only by FireFox browser, while the majority of Internet users use Google and Yandex search engines. This means they simply cannot have access to this text proofreading app. In addition, it can only be used to correct errors of the texts sent via the Internet (emails, messages, publications on social networking sites). However, it cannot be used to correct text written by text editors.

Conclusions. Thus, it is possible to state that the available Armenian electronic proofreading systems practically do not carry out their function, therefore it is necessary to create a new, more complete system. Studying all kinds of Armenian and foreign-language proofreading systems and resources, comparing them with one another, and also identifying existing problems and shortcomings, we propose to develop new principles of the proofreading system for more comprehensive and full proofreading.

To reach this aim the following problems are underlined and liable to the solution:

1) having an electronic database that fully presents the current state of the Armenian language, as it includes not only the main vocabulary of Armenian, but also:

- a number of new words and recently created neologisms
- barbarisms
- proper names (geographical names, personal names, surnames)
- acronyms
- abbreviations

2) providing a complete formal description of the Armenian language, which can later successfully be applied to different computer programs

3) creating "Armenian Electronic Proofreading System", with a free access to the system and available for a wide range of people

4) making it collaborative with foreign language proofing systems which offer proofreading opportunities for the Armenian texts.

REFERENCES

1. Sargsyan Meri. Review of Armenian Electronic Proofreading Systems. Retrieved from http://language.sci.am/sites/default/files/conference/hayeren_elektronayin_srbagman_arhka_hamakar_geri_kknowtyown.pdf
2. <https://www.grammarly.com/>
3. Peterson James. Computer Programs for Detecting and Correcting Spelling Errors (PDF). (Dec 1980), Retrieved 2011-02-18.
4. U.S. Patent 6618697, Method for rule-based correction of spelling and grammar errors, Retrieved from <https://patents.google.com/patent/US6618697>
5. <http://www.hyspell.com/>
6. <http://www.powerspell.am/>
7. Real Estate Cadastre State Committee of the Republic of Armenia/ Geographical Names. http://www.cadastre.am/page/ashxarhagrakan_anvanumner
8. https://www.spellchecker.net/eastern_and_western_spell_checker.html
9. <http://am.translit.cc/>
10. http://www.stars21.com/spelling/armenian_spell_checker.html
11. <https://addons.mozilla.org/en-US/firefox/addon/armenian-spell-checker-diction/>