A Corpus-based Study of Lexicon and Classic Classification of the Main Themes of Children's Poetry Books of the 1390s A.H.

Zeinab Lavafipour

Master of General Linguistics Shiraz University,

zeinab.lavafi@gmail.com (corresponding author)

Jalal Rahimian

Professor of General Linguistics Shiraz University,

jrahimian@rose.shirazu.ac.ir

Introduction

Corpus and corpus studies play an important role in literary studies today. In addition to providing multifaceted studies in one language, corpus and corpus software have enabled quantitative and qualitative analysis and increased the speed and accuracy of research compared to the intuitive methods available in the past. Linguistic corpuses have different types that can include a part of a language or include all of that language in its entirety and are also widely used in literary studies, poetry and drama. To design a corpus, the aim of the creator is very important because the size and volume of the corpus will be determined, and in this regard, the general purposes can better control the data output. In addition to being balanced, the data should also have the potential to be generalized to the language. In many cases, the use of some of the computer software provides a choice of quantitative and

qualitative methods, such as the AntConc tool, which displays homonyms, repetition of words and different aspects of a language.

Methodology and Review of Literature

Children's written texts in many languages have been analyzed using corpus-based methods. Among these corpuses, we can mention CLiC Fiction (Cermakova, 2017), CCB (Göbel & Peetz, 2005), OCC (Lexical computing, 2019) and CLLIP (Thompson & Sealey, 2007). Also, the investigation of poetic themes in Persian children's literature has been undertaken by researchers such as Jorkesh (1975), Zarghani and Abbaspour Noghani (2007), Sangari and Ershadifar (2014), Alimi (2015), Ebrahimi and Salajeghe (2016); however, unlike some countries, there has been no written corpus related to Persian language children's literature in Iran so far, and the lexicons of poetry with a high volume in the nineties A.H. have not been examined in this method either.

Therefore, the aim of the current research is to create a corpus of children's poetry books of the age group of young children (A-B), published by the Institute for the Intellectual Development of Children and Young Adults (as one of the oldest and most reliable publishers of children's books) in the 90s A.H. because in this decade, the data is newer and more diverse, and there have been fewer studies on them than in the past decades. Also, this research has identified and introduced the frequently used and underused words of poems in the corpus and thematic classification of the words and themes of the books in a classical method.

This research, using AntConc software (2019), has investigated vocabulary in a corpus that contains 62 files with 605 poems and 37946 words. The data of group A (years before elementary school) includes 33 written files and more than 16,000 words; the data of group B (early years of elementary school) includes 19 written files and more than 12,000 words; and the data of group A-B includes 10 written files and more than 9,000 words that has been collected through printed and audio books. In this research, after transcribing, naming and saving the data, the free AntConc software was used. This software includes seven abilities but, in this research, the ability of concordance, file view and word list are used more than other features.

Discussion

In all three age groups, 20 most used words were analyzed based on word list and concordance tools, and it was revealed that in all three age groups of preschool and first years of elementary school, the subject of numbers has the largest number of words, because learning numbers and counting is an essential life skill. In each of age groups A and B, separately, the topic of nature has the largest number of words because the child sets foot in nature and the natural environment from birth and human life is a part of nature. Nature is vast and diverse, which leaves poets free to choose words related to the topic. In general, group B has less subject variety than group A. For example, the word God has been used more in group A than in group B because the child's questioning mind needs to be familiar with the non-objective concept of the name of God. The word cat also has been used much more in group A than in group B, which can be due to its domesticity or being heard

continuously in the outside environment and in the home environment as a pet, compared to other animals. Based on 60 frequently used words in group A, B and (A-B), it was found out that most of the commonly used words have short syllables. Poets tend to use simple words according to the age of the child and his or her living environment. Subjects such as food, objects, quality and clothing are in the less frequently used lexicon group. However, there are 10 vocabulary topics in common between the frequently used and rarely used vocabulary. The frequently used vocabulary group often includes more general and comprehensive words, while the less frequently used vocabulary group often includes more specific and less familiar words for the child. For example, a cat is more familiar to a child than a crow; or among the family members, the child has been in contact with his mother and father since birth, rather than with his or her other relatives such as grandfathers and grandmothers.

Due to the metaphorical nature of most of the children's poems, the widely used topics of poetry books have been studied through the classical method, and it was found out that the main contents in all three age groups include naturalistic, entertaining, socio-moral, educational themes, socio-political and religious. Nature is the main theme of children's poetry books in the 90s A.H. because nature is the child's place of play, the source of his inspiration and discoveries. Every day, the child is somehow in contact with nature. He or she is connected with elements such as water and plants in the house and with celestial objects in the outside nature.

Conclusion

The findings of this study indicate that the thematic diversity of words in group A is more than group B and (A and B) and group (A and B) has more thematic diversity than group B. Educational themes in group A and socio-moral themes in group B are in a higher position than other poetic themes.

The results of this study showed that words related to nature were used with a frequency of 21.90%, numbers with 15.32%, family with 13.68% and places with 12.52%, and most of these words are simple and concrete, in such a way that the child has been in contact with since birth. Naturalistic and entertaining themes are among the most widely used poetic contents in the poetry of the 90s A.H. Based on some previous researches in the seventies and eighties A.H., the subject of naturalism has always occupied the first place as the most used subject. That the religious and socio-political issues are in the next places, which is slightly different from the results of this research, can be due to the use of data from an older age group in previous researches. In general, poets have shown more interest in using simple and tangible words with a positive and happy semantic load appropriate to the age of the child.

Keywords: AntConc Software, children's poetry, Corpus-based studies, lexicon, themes

References

Alimi, M. (2015). Study and important elements of the content of children's poetry in the 80s A.H. *Adabiyate Farsi* [*Persian*

Literature] *(Islamic Azad University of Mashhad), 47*(11), 120-140. [in Persian]

Alipour, M. (2015). *Pazhouheshi dar shere koudak* [*A research on children's poetry*] (8th ed). Tehran: Tirgan. [in Persian]

Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University, Retrieved December 15, 2020 from: http://www.laurenceanthony.net/software/antconc/.

Asi, M. (2000). From corpus language to corpus linguistics. *Allameh Tabatabaei University, 91*, 488-495. [in Persian]

Cermakova, A. (2017, November 28). The GLARE 19th Century Children's Literature Corpus in CLiC. Retrieved from https://blog.bham.ac.uk/clic-dickens/2017/11/28/the-glare-19th-century-childrens-literature-corpus-in-clic/

Davari, Z. (2018). *Yek ghach hendevane* [A watermelon wedge] (3rd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Ebrahimi, J. (2017). *Chistam man* [Who am I] (4th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Ebrahimi, J., Maleki, B., Shabannejad, A., Shabani, A., Ghasemnia, Sh., Niktalab, B., & Rahmandoost, M. (2016). *Arousie Arousi* [Wedding, wedding] (5th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Ebrahimi, J., Shabannejad, A., Shabani, A., Ghasemnia, Sh., & Rahmandoost, M. (2017). *Taranehaye khabo khial* [Dream and fantasy chansons] (5th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Ebrahimi, J., Shabannejad, A., Shabani, A., Ghasemnia, Sh., Niktalab, B., & Rahmandoost, M. (2017). *Taranehaye babaie* [Daddy chansons] (5th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Ebrahimi, Sh., & Salajeghe, P. (2016). Analysis and comparison of naturalistic themes in Iranian and English modern children's poems. *Advances in Language and Literary Studies, 7*(4). doi:10.7575/aiac.alls.v.7n.4p.181

Eshaghi, A. (2015). *Dane va hendevane* [Seeds and watermelon]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Eslami, M. (2014). *Damadi-e agha kalaghe* [Mr. Crow's wedding]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Eslami, M. (2015). *Yoohoo haha seda miad* [Yoohoo haha there is a sound]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Evans, D. (2004). Corpus building and investigation for the humanities: An on-line information pack about corpus investigation techniques for the Humanities. *Linguistics*, 15–16. https://doi.org/10.1109/SocialCom.2010.106

Farhadi, P. (2020). *Arous kie, doomad kie* [Who is the bride, who is the groom] (2nd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Ghasemnia, Sh. (2015). *Man o maman o madrese* [Me and mom and school]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Ghasemnia, Sh., Shabani, A., Shabannejad, A., Ebrahimi, J., & Niktalab, B. (2018). *Sher-e doa* [Prayer chansons]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Göbel, T., & Peetz, M. H. (2005). CCB: A corpus of children's books. Retrieved from http://peetz-intelligence.com/http://peetz-intelligence.com/

GoliFarhoud, S. (2019). *Negahi be adabiyate koudak* [A look into children's literature]. Khorasan Razavi: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Hartman, J. (1932). The place of poetry in children's literature. *The Elementary English Review, 9*(1), 17-19. Retrieved December 15, 2020, from http://www.jstor.org/stable/41381465

Hashempour, M. (2015a). *Mioo mioo tashakor* [Meow meow thank you] (2nd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Hashempour, M. (2015b). *Vay agar dom dashtam* [Oh if I had a tail]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Hashempour, M. (2018). *Yek ghouri por az ghour* [A teapot full of croak] (3rd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Hashempour, M. (2020). *Badbadakam be abrha tekie dade* [My kite rests on the clouds] (3rd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Hassani, Kh. (2016). *Dota sibe shirin* [Two sweet apples]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Hassani, Kh. (2016). *Mah mige dali doushe* [Moon says peekaboo] (3rd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Jorkesh, Gh. (1975). Poetic themes in the second elementary book: a look on children's poetry in Iran. *Negin,123*(11), 15-22. [in Persian]

Keshavarz, N. (2015a). *Payiz khanom* [Ms. Autumn] (5th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Keshavarz, N. (2015b). *Par-e zard-e ghanari* [Canary's yellow feathers] (5th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Keshavarz, N. (2016a). *Fereshte-i az asemoon* [An Angel from the sky] (3th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Keshavarz, N. (2016b). *Kalagh-e zard* [Yellow crow]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Lexical Computing CZ s.r.o. (2019, April 29). Oxford Children's Corpus. Retrieved from https://www.sketchengine.eu/oxford-childrens-corpus/

Mahouti, M. (2015). *Bache pirahan* [Baby shirt]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Mahouti, M. (2016). *Bourim bazar* [Bourim bazaar]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. New York: Cambridge University Press.

Mohammadi Rafei, Sh. (2005). The importance of basic vocabulary in children's poetry. *Ketabe Mah Kudako Nojavan* [*Moon Book, Iranian Children and Adult Youth's Book Review and Information*], *101-103*(9), 175-180. [in Persian]

Moharami, R., Asadolahi, Kh., PourKhales, Sh., & Foroughirad, P. (2017). Studying the themes of Mahmoud Kianoosh's two collections of children's poetry with the audience approach. *Faslnameye Pazhouheshhaye Adabi va Belaghi* [*Quarterly Journal of Literary and Rhetorical Research*], *5,2-1*(18), 54-70. [in Persian]

Mousavi, M., & Keshavarz, N. (2014). *Ghad-e sar-e souzan* [As big as a needle head]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Niktalab, B. (2018). *Khaharam parande shod* [My sister became a bird] (2nd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Niktalab, B., Ghasemnia, Sh., Shabani, A., Ebrahimi, J., Rahmandoost, M., & Shabannejad, A. (2018). *Az asemoon ta inja* [From the sky to here] (5th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Niktalab, B., Ghasemnia, Sh., Shabani, A., Ebrahimi, J., Rahmandoost, M., & Shabannejad, A. (2017). *Bazi bazi bazi* [Play play play] (4th

ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Nilsson Björkenstam, K. (2013). *What is a corpus and why are corpora important tools?* 5. www.antlab.sci.waseda.ac.jp/

O'Keeffe, A., & McCarthy, M. (2010). *The Routledge handbook of Corpus Linguistics*. New York: Routledge.

Ostadzadeh, Z. (2012). Content analysis of Parvin Dolatabadi's children's poems, *Bar ghayeghe abrha* va *Gole badam*. *Motaleate Adabiyate Koudak* [*Children's Literature Studies*], *3*(2), 1-24. [in Persian]

Rahmandoost, M. (2016). *Sagi bood jangali bood* [There was a dog, there was a jungle] (7th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Rahmandoost, M. (2017a). *Baba amad nan avard* [Dad came and brought bread] (7th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Rahmandoost, M. (2017b). *Taranehaye navazesh* [Caressing chansons] (6th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Rahmandoost, M. (2020a). *Oun chie ke, oun kie ke* [What is that, who is that] (5th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Rahmandoost, M. (2020b). *Khale jan Dido va chand ta gerdoo* [Dear Aunt Didoo and some walnuts] (2nd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Rahmandoost, M. (2020c). *Kane-ye khale* [Aunt's house] (5th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Rahmandoost, M., & Hemati Ahoui, A. (2020). *Bazi ba angosht-ha* [Games with fingers] (8th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabani, A. (1995). Form and structure of children's poetry. *Research Journal of Children and Literature* [*Pazuheshnameh Adabiyat Koodak va Nojavan*]*, 2*(1), 22-29.

Shabani, A. (1995). Ma gol-haye khandanim, farzandane Iranim: A research on children's poetry in Iran. *Goharan, 6*(1), 125-142. [in Persian]

Shabani, A. (2015). *Rangin kamane koudakan* [Children's rainbow] (3rd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabani, A. (2016a). *Tab tab tab-bazi* [Swing swing swinging]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabani, A. (2016b). *Mah-e now, negah-e now* [New moon, new look] (2nd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabani, A. (2016c). *Har kasi kari dare* [Everyone has a job] (4th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabannejad, A. (2013a). *Zanbour o sibe ghermez* [Bee and red apple] (4th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabannejad, A. (2013b). *Gorgam-o galleh mibaram* [I'm the wolf and take the sheep] (3rd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabannejad, A. (2015a). *Boodo boodo bood 1* [Was and Was 1]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabannejad, A. (2015b). *Boodo boodo bood 2* [Was and Was 2]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabannejad, A. (2017a). *Abr umad bad umad* [Cloud came, wind came] (9th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabannejad, A. (2017b). *Arang arang begoo che rang* [Arang arang, tell the color] (4th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabannejad, A. (2017c). *Chatri az golbarg-ha* [An umbrella of petals]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabannejad, A. (2017d). *Kalaghe kojast* [Where is the crow] (6th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabannejad, A. (2020a). *Man va khabe jangal* [Me and the dream of the jungle] (4th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabannejad, A. (2020b). *Hajastam o vajastam* [I searched and searched] (2nd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shabannejad, A., & Parsanejad, K. (2016). *Kafsh-e touri gol-e bolouri* [Lace shoes, crystal flower] (2nd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shafiei, Sh. (2020). *Abre ghelghelaki* [Tickle cloud] (2nd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shamani, T. (2016). *Ghech ghech* (2nd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shamani, T. (2018). *Dom-e gorbe-haro bekesh* [Pull the cats' tails] (3rd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Shirazi, P. (2020). *To mahi-o man adam* [You are a fish, I am a human] (3rd ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Szudarski, P. (2018). *Corpus Linguistics for Vocabulary*. London: Routledge.

Taghi, Sh. (2015). *Chahar fasl* [Four Seasons] (10th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Thompson, P., & Sealey, A. (2007). Through children's eyes? *International Journal of Corpus Linguistics, 12*(1), 1-23. doi:10.1075/ijcl.12.1.03tho

Vatani, I. (2014). *Sar-e kouhi residam* [I reached the top of the mountain] (5th ed). Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

VazifeShenas, Sh. (2015). *Sar sari das dasi*. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

VazifeShenas, Sh. (2016a). *Sousmar* [crocodile]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

VazifeShenas, Sh. (2016b). *Ghourbaghe* [Frog]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

VazifeShenas, Sh. (2016c). *Lakposht* [Turtle]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

VazifeShenas, Sh. (2016d). *Hezarpa* [Centipede]. Tehran: Institute for the Intellectual Development of Children and Young Adults. [in Persian]

Zarghani, S., & Abaspour, M. (2007). Content analysis of children's poetry in the seventies (age groups A, B and C). *Majaleye Motaleate Adabi* [*Journal of Literary Studies*]*, 159*(40), 167-192. [in Persian]

Zen, E. L. (2018). Corpus-driven analysis on the language of children's literature. *Proceedings of the 1st International Conference on Recent Innovations*. doi:10.5220/0009912700170022