

LINGVISTIKA/LINGUISTICS

Using Corpora to Obtain Social and Cultural Information: A Case Study of *America*

Jūratė Vaičenonienė

Abstract. The subject of the present article has been triggered by the recent rush to apply corpora to a variety of researches. Huge electronic texts are employed to obtain not only data for linguistic investigations, but their application has been extended to other branches of science as well. The present analysis of the proper noun *America* is an attempt to use abundant linguistic data to determine social and cultural factors related to this particular geographical name. It also aims at distinguishing the attitudinal meaning of *America*. The research is based on the collocations of the word under analysis derived from the Lithuanian press that is a constituent part of the Lithuanian corpus.

Traditionally proper nouns are treated as linguistic units that signify specific or unique objects in the world. They are contrasted to common words as being neutral and functioning merely as signifiers of places, organisations or people. In this sense, *America* (here and further italicised when it is treated as a lexeme under investigation, not a reference to a place) is an ordinary place-name with no other meanings. However, the present analysis is an attempt to investigate the proper noun *America* in the Lithuanian press in order to reveal its attitudinal meaning as well as social and cultural associations. By examining the collocates of the noun the general attitudes to this country can be revealed. As the analysis of the connotation of the noun *Kaunas* carried out by R. Marcinkevičienė shows, proper nouns just as common words have meanings (Marcinkevičienė, 2000: 137-147). She demonstrates that proper nouns are used not just to refer to places, people or other unique objects, but they also invoke certain associations and are used in a general sense. Furthermore, a linguistic study of connotation can lead to certain social and cultural generalisations.

The area of onomastics is a long-established field with a range of topics analysed. Franz Boas in his analysis of the geographical names of Kwakiutl Indians rightly points out that “geographical names, being an expression of the mental character of each people and each period, reflect their cultural life and the line of development belonging to each cultural area” (Boas, 1964: 171). He traces cultural aspects by examining the etymology of the proper nouns in that particular tribe and by analysing the inherent meanings of the nouns. Thus traditionally the motivation of a particular proper noun is at the focus of toponymic analyses. Ethno linguistics treats onomastics from the same point of view, i.e. they examine the meaning of proper nouns by tracing their history as far back as their origins (Gudavičius, 2000: 104-108). However, the present investigation is just the reverse of the studies like Boas’s. We are interested not in the origins of *America*, but in its present connotation and associations. In other words, not its past meaning is at the focus, but rather the notions that

the Lithuanian language users ascribe to the proper noun in the Lithuanian press nowadays.

The significance of the attitudinal meaning of proper nouns arises from the general standpoint of John Sinclair, who reveals the importance of semantic prosody “in the study of lexical grammar” (Sinclair, 2001: 198). He notices the fact that words often have attitudinal meanings, e.g. “*scrawny* is unpleasantly thin and bony” (Sinclair, 2001: 199). Thus certain words tend to collocate with specific groups of words and therefore they gain some additional connotative aspect of meaning that should be represented in dictionaries. Although he refers only to common words, the same principle can be applied to proper nouns to a certain extent. The necessity to comment on the connotation of all nouns is proposed by Rosamund Moon who participated in the COBUILD project. She illustrates the point with the metonymic case of *Whitehall* when it signifies the British government (Moon, 1987: 99). Similarly, *America* in a number of cases evokes specific associations and gains connotative meaning, as will be demonstrated in the article.

The source of the data for the research is the Lithuanian corpus of a variety of texts that contains 60 million words or tokens. For the specific purposes of the analysis only several parts of it were used, i.e. the local press, central periodicals, specialised periodicals, popular periodicals and the central press. The proportions of the entire corpus are illustrated in Figure 1. As can be seen in the figure, the press makes up the total of 72 %.

The information that was retrieved from the corpus was hardly manageable. *America* is definitely a popular topic in the Lithuanian press. There are over 11 thousand occurrences of *America* in the press. Hence to limit the amount of the data, the first 600 collocates were analysed to a greater detail. The frequency of these collocates is from 300 to 20 times in the multi-million corpus. Thus, the collocates under the further analysis are of an outstandingly high frequency. They can be divided into the following major groups:

- specific geographical place names
- other countries that appear in association with America
- words related to migration
- words signifying different life spheres

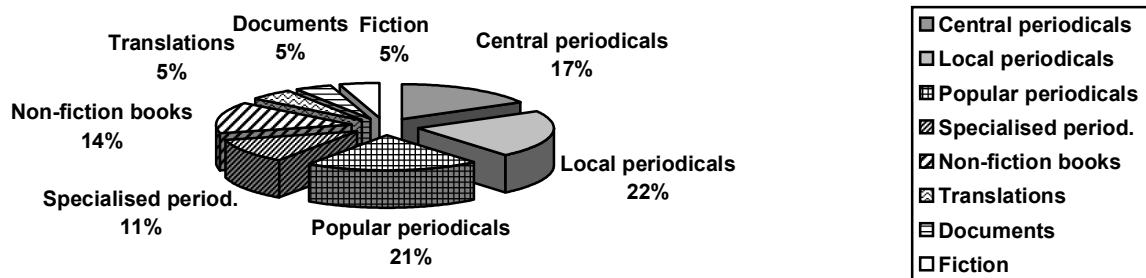


Fig. 1. The structure of the Lithuanian corpus

There is a group of collocates that are too various to be grouped under separate headings and will be overviewed separately.

First of all, among the first 600 collocates of *America*, the proper name under analysis is used clearly as a geographical name, i.e. South America (298 times), North America (255 times), Latin America (186 times), Western America (98 times), Eastern America (73 times) and Central America (25 times). On the one hand, these are just geographical terms and appear to be the most neutral collocates of America. On the other hand, they suggest the focus of the Lithuanian press on certain regions. Similarly, in relation to America, certain cities are mentioned more frequently than others, i.e. Chicago (75), Washington (49), Atlanta (28) and Los Angeles (24). The high frequency of the latter cities can be rather easily explained. Chicago is the main focus of interest for the Lithuanian press because it has one of the greatest Lithuanian communities in America. Washington, being the capital of the country, is naturally discussed to a wider extent than other places. It especially frequently appears in relation to the White House or the official visits of Lithuanian politicians and diplomats to the capital of America. The high frequency of the collocate ‘Atlanta’ reflects the interest in the Olympic Games that were held there in 1996. Thus the frequency list of collocates can be a very informative source of such aspects as cultural, political or other contacts with another country and major events in that country.

The countries or regions that are mentioned in relation to America 20 times or more are Africa, Russia, Europe, Germany, Canada, Asia, Australia, England, France and Japan. The countries have been listed in the order of their frequency. The above-mentioned countries are either referred to as contrasting with America (especially Russia and Europe) or they evoke certain economic associations, i.e. Africa is a poor country that is being helped by America. In general, America is treated as a country one can expect to obtain help from. The latter tendency becomes evident from the collocate ‘help’ used in different forms which all in all make up the frequency of 40 times. Canada is a geographically related country, whereas Asia is the territory of America’s economic interests. Besides, Canada, Germany, England, France and Japan are major countries that are equal to America because they all are

well-developed and economically as well as politically powerful countries.

Surprisingly enough, such problems as different racial groups are scarcely mentioned in relation to *America*, i.e. the frequency of ‘baltieji’ (the white) in the corpus is only 16 occurrences and the frequency of Indians is 40. ‘Juodieji’ (the black) are not mentioned at all. Instead, the offensive noun ‘negro’ is mentioned 4 times.

Feministic tendencies in America, in contrast, have a significant place in the Lithuanian press. The collocate ‘moterų’ appears 32 times. The sentences they appear in mainly reflect women’s fight for their rights, e.g. *kova už moterų politines teises* (‘the fight for women’s political rights’). One more interesting detail becomes evident from the samples of the latter collocate, i.e. *Amerikoje nėra gražių moterų* (‘There are no beautiful women in America’). The example reflects the common conviction or even a cultural stereotype existing in Lithuania that women in America are not beautiful.

The spheres of life that appear in relation to America are various, but the most noticeable one is work. It is used in different word forms, i.e. as a verb ‘to work’, which is used in different tenses, as well as a noun that is used in different cases. The frequency of the collocate is represented in Table 1.

Table 1. Frequency of the collocate ‘work’

No.	Word form	Frequency number
128	darbo	86
185	darbą	61
251	dirbo	47
257	dirbti	46
486	darbas	25
Total: 265		

Consequently, the issue of money is among the most frequent collocates of America, i.e. *dolerių* (‘dollars’) (85) and *pinigų* (‘money’) (52). The words mainly appear in the context that Lithuanians go there with the intention to earn it. Furthermore, the issue of emigration and travelling is especially frequent as the list below reveals:

Įvyko gyventi (‘left to live’)	57
Gyvenantys (‘living’)	24
Gyvendamas (‘when living’)	24

Gyvenantis ('living')	23
Išivijojos ('of emigrants')	37
Kelionės ('trips' or 'of a trip')	31
Kelionę ('trip')	27
Atvykęs ('having come')	29
Atvyko ('came')	26
Grįžti ('return')	24
Išvažiavo ('left')	23

In relation to the migration issue, it should be noted that two most frequent collocates of America are two prepositions, i.e. 'į' (1.749) and 'iš' (975). The prepositions suggest the intense contact and exchange of people, culture and ideas between the two countries. The most popular season for going to America is obviously, summer as the phrase 'vasarą vykti' ('to go in summer') occurs 29 times in the corpus part under investigation. As the corpus examples suggest, the main reasons for going there in summer are to study, to participate in sports competitions and to work on a cruise ship. The contexts of the collocate 'ieškoti' ('to look for') provide some more reasons for travelling to America, i.e. ieškoti laimės (3), darbo (4), genčių (1), verslo idėjų (1), rėmėjų (1), gabių mokinių (1). Thus Lithuanians seek happiness, work, tribes, business ideas, sponsors and talented students in America.

The samples of the collocates of the adverbs 'gerai' ('well') and 'geriau' ('better') can serve as an explanation why America is a land of dreams as a very frequent collocation 'svajonių šalis' ('land of dreams') (13) suggests. Thus in America

'gerai jautiesi' ('you feel better'), 'gerai uždirba' ('one can earn a lot'), 'gerai apsirengę' ('are well-dressed'), 'viskas gerai' ('all is well'), 'labai gerai sekėsi' ('was very successful'), 'norime gyventi gerai – kaip amerikoniai' ('we want to live just as well as Americans'); 'Amerikoj daug geriau' ('It's much better in America'), 'moteriai menininkei geriau pasilikti Amerikoje' ('for a woman artist it is better to stay in America').

However, the evaluation of the American lifestyle is rather contradictory, e.g.. *Lietuvoj geriau negu Amerikoje* ('It's better in Lithuania than in America'). The adjective 'sunku' ('difficult') (37) is even more informative. Things that might be complicated in America are as follows:

susrasti naujų draugų labai sunku ('it's very difficult to find new friends')
gauti gerą darbą labai sunku ('it's very difficult to get a good job')
turėti savo biznį labai sunku ('it's very difficult to have one's own business')
išlikti Amerikoje labai sunku ('it's very difficult to survive in America')
Amerikoj būt laimingam sunku ('it's difficult to be happy in America')

In addition to the problem of work and money, the spheres of culture, education, politics and sports appear to be important in the Lithuanian press in relation to America:

Table 2. Life spheres that appear in relation to America

Sphere	No.	Collocate	Frequency
Culture	190	kultūros	60
	661	kultūra	19
	267	meno	44
	498	teatro	25
	570	filmų	22
	1023	literatūros	13
Sports	305	krepšinio	38
	929	krepšininkai	14
	1013	krepšininkų	13
	320	sporto	37
Studies	613	studijuoti	21
	673	studijų	19
	814	studijas	16
	1049	studentų	13
	1050	studijavo	13
Politics/ Diplomacy	347	Seimo	35
	401	politikos	30

Thus culture is the second most frequent sphere after work that collocates with *America*. The collocates include a variety of cultural phenomena such as art, theatre, films and literature. If we count the frequency of all the nouns that relate to culture, the overall frequency is as high as 183. The total number of the occurrences of the collocates referring to sports is 102. As it could be expected, 'basketball' and 'basketball players' are collocates of a very high frequency. This fact reflects both America's and Lithuania's interest in this particular branch of sports. Studies in America can be distinguished as one more significant issue that appears in relation to America. The derivatives of the word 'studijuoti' ('to study') total 82 occurrences. Finally, politics and diplomacy are among the most frequent subjects that co-occur with *America*.

It is interesting to notice that *America*, as it was mentioned at the beginning, has become a common noun to refer to something new that can be discovered or a land of dreams. The frequency of such cases amounts to 17. When it is used in this sense, it is not capitalised:

- 1 vėl beveik visi išsilakstė po visokias amerikąs ('almost everyone has run away to different americas')
- 2 garbingos ligoninės neiškeitė į jokias amerikąs ('they didn't change the respectable hospital to any americas')
- 3 išblaškė daugelį kūrėjų po tolimiausias amerikąs ('many artists were scattered in the far away americas')
- 4 "žengia į Europą" ir atranda savo amerikąs ('go to Europe and discover their own americas')
- 5 lietuviai bėga į visokius taivanius, amerikąs ('Lithuanians run to different taiwans, americas')
- 6 Ne vienas pranešėjas "atradinėjo amerikąs" sakydamas, kad mokslas ("quite a few presenters 'discovered americas' by saying that science").

These are the cases when the proper name is used in the plural. It should be noted that this generic meaning is not provided in dictionaries although it is used in language relatively frequently. *Collins COBUILD English Dictionary* distinguishes two senses of *America*:

- 1-the continent containing North, South and Central America;
- 2-the shorter name for the United States of America. (COBUILD, 1995: 52)

However, general associations have influenced the meaning of the proper noun so much that it has gained the additional meaning of, firstly, a foreign country that is a very promising and common destination of emigration. Secondly, because of the historic circumstances it has gained the meaning of something that is new or unknown and can be discovered.

Finally, we should briefly touch upon some cases that are very noticeable in the list of collocates, but that do not belong to any of the groups mentioned above. First of all, a very common collocate of *America* is 'pirtyje' ('in the bath'). Its number in the entire frequency list of collocates is 115 and its frequency is 98 occurrences. It alludes to the play 'Amerika pirtyje' by Keturakis (26). One more reality of our life that is evident from the corpus examples is the radio 'Amerikos balsas' as the collocate 'balsas' (35) suggests. In addition, a very frequent collocate of *America* is 'vaikai' ('children') (42). The children that are referred to in the corpus were either born there or have found happiness there after leaving Lithuania, e.g.

(randa prieglobstį) nelaimingi vaikai iš Lietuvos ('unhappy children from Lithuania find refuge')
Važiukite, vaikai, Amerikon ('go, children, to America').

A surprising collocate of *America* is 'cats' that is not among the first 600 hundred collocates but nevertheless is of a rather high frequency. It appears 20 times in different cases. It is mentioned in relation to what species of cats are brought from *America*, originated there or are not found there.

This brief overview of the most suggestive collocates of the proper noun *America* has revealed that corpus samples can be a very informative source to elicit the attitudinal meaning of a country. *America* appears to be associated with strong and powerful countries. It is treated as a tempting destination for many Lithuanians to earn money and to find happiness although certain difficulties of living there are referred to. It is often depicted as a land of dreams, but it also appears to have a number

of disadvantages. Certain places are emphasised more than others in the Lithuanian press. The latter tendency leads to some generalisation about possible contacts between the two countries as well as events that are of specific importance for us. The notions that *America* is associated with influence the change of the proper noun into a common noun that no longer refers to a geographical region, but rather to its general connotation.

To conclude, it can be said that language studies have undergone great changes thanks to the advance of modern technologies in our society. Huge electronic texts enable researches to carry out more exhaustive language investigations. Moreover, we have access to abundant data that can be applied to studying social and cultural changes. Proper names, if studied using corpora, can serve to disclose their attitudinal meaning as well as cultural, social and political aspects. Collocation should be introduced in language teaching because if proper nouns, which are considered to be neutral, have connotations, common nouns are even more inclined to have associative meanings. The present article is based on a synchronic approach, but if it were possible to use corpora of different dates, changes in social and cultural attitudes to a particular country could be distinguished.

References

1. Boas, F. (1964) On Geographical Names of the Kwakiutl. *Language in Culture and Society*. New York, Evanston, London: Harper and Row, Publishers.
2. Collins Cobuild English Dictionary. (1995) (1st ed.) London: Harper Collins Publishers.
3. Gudavičius, A. (2000) *Etnolingvistika*. Šiauliai.
4. Marcinkevičienė, R. (2000) Kauno vardo konotacija ("Vardažodžių semantika"). *Kauno istorijos metraštis*. Kaunas: VDU leidykla. P.137-147.
5. Moon, R. (1987) *The Analysis of Meaning. Looking Up: An Account of the COBUILD Project*. London and Glasgow: Collins ELT. Pp.86-103.
6. Sinclair, J. (2001) *Lexical Grammar*. *Darbai ir dienos*. Pp.191-204. Kaunas: VDU leidykla.

Jūratė Vaičenonienė

Kompiuteriniai tekstynai kaip duomenų bazė kultūriniais ir socialiniams aspektams tirti (tikrinio daiktavardžio *Amerika* analizė)

Santrauka

Šis straipsnis yra tikrinio daiktavardžio *Amerika* trumpa analizė, paremta kompiuteriniais tekstynais. Tyrimui buvo panaudota Lietuvos spauda, sudaranti 72 % viso tekstyno. 600 dažniausių kolokatų analizė parodė, jog tikriniai daiktavardžiai kaip ir bendriniai žodžiai konotuoja tam tikrus dalykus ir sukelia tam tikras asociacijas. Žodžio *Amerika* vartoseną su tam tikrais kolokatais išryškina Lietuvos žiniasklaidos dėmesį tam tikroms jos sritims ir miestams. Sugrupavus dažniausiai pasikartojančius žodžius išryškėja gyvenimo sferos, susijusios su *Amerika*, t.y. darbas, kultūra, sportas, studijos, politika ir diplomatija. *Amerika* dažnai įsivaizduojama kaip svajonių šalis, nors dažnai gyvenimas šioje šalyje vertinamas gana kontraversiškai. Šie ir kiti aspektai įtakoja tiriamąjį vietovardį taip, kad šis ganėtinai dažnai yra vartojamas kaip bendrinis daiktavardis daugiskaitoje ir yra rašomas mažąja raide. Tokiais atvejais *Amerika* įgyja dvi reikšmes. Pirmą, dėl asociacijų su ekonominiu klestėjimu ji žymi šalį, kur matomos geresnio gyvenimo perspektyvos. Antra, dėl istorinių aplinkybių šios šalies pavadinimas yra įgijęs naujo dalyko, kurį galima atrasti, reikšmę. Tad tekstynas ne tik pakeitė lingvistinius kalbos tyrimus, bet ir gali būti sėkmingai taikomas kitose srityse, pavyzdžiui, socialinėms ir kultūrinėms tendencijoms bei pokyčiams tirti. Kolokacija turėtų tapti neatsiejama kalbų mokymo dalimi ne tik dėl to, kad atskleidžia žodžių vartosenos ypatumus, bet ir dėl to, kad išryškina minėtų ekstralingvistinius aspektus bei konotacinius tikrinių ir juo labiau bendrinių žodžių atspalvius.

Straipsnis įteiktas 2001 04
Parengtas publikuoti 2001 06

The author

Jūratė Vaičenonienė, M.A., assistant lecturer.

Academic interests: corpus linguistics and its applications to lexicography, cognitive linguistics.

Address: Vytautas Magnus University, Donelaičio str. 52, LT-3000 Kaunas, Lithuania. *E-mail:* v_jurate@hotmail.com.