Agnieszka Dryjańska (Warszawa)*
ORCID: 0000-0003-1649-8408

Patrimoine vs. *dziedzictwo* – interculturality in French language teaching

The subject of this paper is a corpus analysis of *patrimoine* in contrastive perspective with its Polish equivalent *dziedzictwo* within the framework of the intercultural approach in French language teaching. Its purpose will be to reveal the semantic differences and similarities of these two words in terms of the results provided by the frequency and collocation analysis based on the Polish National Corpus, the French Corpus Frantext and Corpora Collection of Leipzig University. The study showed that one of the strongest and most frequent collocations, indicated by different collocation measures, for both Polish and French, is *cultural heritage* (Fr. patrimoine culturel, Pl. dziedzictwo kulturowe). Typical Polish collocations are *national heritage* and *Christian heritage*, while in French these are *patrimoine artistique* et *patrimoine touristique*.

Keywords: heritage, corpus analysis, collocation, French language teaching

Patrimoine vs *dziedzictwo* – interkulturowość w nauczaniu języka francuskiego

Przedmiotem artykułu jest analiza francuskiego słowa *patrimoine* w perspektywie kontrastywnej polskiego słowa *dziedzictwo* w kontekście podejścia interkulturowego w nauczaniu języka francuskiego. Jej celem będzie wskazanie różnic i podobieństw semantycznych tych dwóch słów w oparciu o rezultaty badania korpusowego, obejmującego analizę frekwencyjną i kolokacyjną w Narodowym Korpusie Języka Polskiego, francuskim korpusie Frantext oraz Corpora Collection of Leipzig University. Badanie wykazało, że wspólną (częstą i jedną z najsilniejszych) wskazywaną przez różne miary kolokacją jest dla języka polskiego i francuskiego jest *dziedzictwo kulturowe* (fr. patrimoine culturel). Kolokacje typowe dla języka polskiego to *dziedzictwo narodowe* i *dziedzictwo chrześcijańskie*, natomiast dla języka francuskiego są to *patrimoine artistique* i *patrimoine touristique*.

Słowa kluczowe: dziedzictwo, analiza korpusowa, kolokacja, nauczanie języka francuskiego

* dr Agnieszka Dryjańska, Uniwersytet Warszawski, Instytut Romanistyki, ul. Dobra 55, 00-312 Warszawa, e-mail: a.dryjanska@uw.edu.pl

Patrimoine vs. *dziedzictwo* – Interkulturalität im Französischunterricht

Gegenstand des Artikels ist die Analyse des französischen Begriffs *patrimoine* und sein Vergleich mit dem polnischen Wort *dziedzictwo* vor dem Hintergrund des interkulturellen Ansatzes im Französischunterricht. Das Ziel des Analyse ist es, die semantischen Unterschiede und Ähnlichkeiten zwischen den beiden Wörtern auf der Grundlage der Ergebnisse einer Korpusanalyse aufzuzeigen, einschließlich der Frequenz- und Kollokationsanalyse im Nationalen Polnischen Sprachkorps, im Französischen Frantextkorpus und im Kollokationskorpus der Universität Leipzig. Die Studie zeigte, dass der Ausdruck *kulturelles Erbe* (fr. *patrimoine culturel*) im Polnischen und Französischen eine gemeinsame sowie besonders frequente und starke Kollokationen ist. Als typisch im Polnischen gelten Kollokationen wie *dziedzictwo narodowe* i *dziedzictwo chrześcijańskie* [nationales und christliches Erbe], während im Französischen *patrimoine artistique* und *patrimoine touristique* dominieren.

Schlüsselwörter: Erbe, Korpusanalyse, Kollokation, Französischunterricht

1. Introduction

The centenary of the Polish independence has become an opportunity to reflect again on the meaning and place of our national heritage, both in a historical and contemporary perspective. Various types of cultural events organized on this occasion resulted in an increase in the popularity of the Polish word *dziedzictwo* (Eng. heritage). These types of events, important for culture, also have an impact on education, including foreign language teaching. Consequently, the subject of my study is a corpus analysis of *patrimoine* in the contrastive perspective with its Polish equivalent *dziedzictwo* in order to integrate its elements in foreign language teaching within the framework of the intercultural approach.

Although my analysis was inspired by the analysis of the cultural key word *homeland* (Fr. patrie, Pl. ojczyzna) conducted by Polish ethnolinguists (Bartmiński 1993; Kłoskowska 1993; Abramowicz 1993 and Wierzbicka 1997), it differs in many aspects. Firstly, I have chosen the key word *patrimoine* for didactic reasons. It appears in students' books for French language teaching, often in the context of tourism, as we can read in the coursebook *En Action 2* (level A2):

> The French start to replace 'beach-sunbathing-idleness tourism' with 'intelligent tourism', in order to discover the rich historical and cultural <u>heritage</u> of their regions (En Action 2, p.40).[1]

---

[1] « Les Français commencent à remplacer le tourisme « plage-bronzage-farniente » par le « tourisme intelligent », afin de découvrir le riche <u>patrimoine</u> historique et culturel de leurs régions ».

What is interesting in this sentence is the local dimension of heritage, a link between heritage and regions, not between heritage and the nation, nor the homeland, which would be a more intuitive choice for Poles. The analysis of even such a short sentence can be an occasion to reflect on the differences and similarities between the Polish and French understanding of *heritage*. From my perspective, such a reflection, founded on reliable data, contributes to the development of an intercultural approach.

Secondly, I based my research on the contrastive approach, crucial for the interculturality in foreign language teaching, not present for instance in the analyses of *patrie* by Kłoskowska (1993) or Abramowicz (1993). The latter notes the importance of a local factor in the definitions of *patrie* in French dictionaries, but I find similar aspects in Bartmiński's analysis of *ojczyzna*. These analyses, as they are not anchored in the contrastive perspective, do not permit conclusions about whether there are any differences in the meaning of the words *ojczyzna* and *patrie*.

Finally, I based my research on a corpus analysis that provides more reliable data in terms of the quantity and variety of the sources, especially when big corpora are concerned. Moreover, corpora contain data and tools more accessible to foreign language students, who usually are not sufficiently familiarized with a foreign culture, literature, history, and other fields rigorously explored in ethnolinguistic research. That is why I propose a corpus driven approach that, on the one hand, reveals the differences in meaning of *patrimoine* and *dziedzictwo* in terms of the results provided by a frequency analysis, and an analysis of collocations based on the Polish National Corpus, the French Corpus Frantext and Corpora Collection of Leipzig University. On the other hand, it also permits a conclusion to be made about what technical aspects of a corpus analysis should be taken into account if we want to develop a method that could be applied by learners of a foreign language.

2. Intercultural personality and its sources

The intercultural perspective in foreign language teaching lets the learners, immersed in their own culture and language, commence discovering other cultures through learning a new language. This, as a result, allows them to deepen the comprehension of their own culture. The process impacts their attitude to languages and cultures in general.

'Interculturality' in foreign language teaching can be developed in many ways. One of them is the development of lexical competence (CEFR 2001) with a special emphasis on cultural key words. *Lexiculture* was the French term proposed by R. Galisson (1998) to emphasize a close relationship between lexis

and culture to enable learners to understand that words in different languages, especially cultural key words, are not easily replaceable labels, but that words such as *freedom* or *homeland* are "a society's cultural artifacts" (Wierzbicka, 1997 on line). R. Galisson, in the context of foreign language teaching, talks about words with 'added value' (Fr. charge culturelle partagée) (Galisson 1987 in Skibińska, 2003). They convey mental concepts that significantly differ from one language to another. The 'added value' of words should be taken into consideration in the process of teaching foreign languages as the consequently formed lexical competence "serves interpersonal communication" (Skibińska 2003: 6).

Intercultural personality and competence are considered nowadays as key terms in foreign language teaching. The first one is formed by "both attitudes and awareness of things" and is in itself an "important educational goal in its own right" (CEFR 2001). Its development is possible thanks to intercultural competence, defined as a set of competences, which are developed when a foreign culture is superimposed on that of the learner of the foreign language and leads to a change of his or her attitude[2] (Dryjańska 2016). The process of developing the intercultural personality is therefore triggered by one's contact with a foreign language. The learner confronted with cultural differences begins to question both *otherness* and his or her own linguistic and cultural habits. It is the moment of birth of his or her intercultural personality. This mechanism can take place in a language class if the teacher considers the language as a vector of culture and tries to transmit this vision to learners by an appropriately chosen methodology of vocabulary teaching. Such concept of language is not new. John Locke noted that we could "observe a great store of words in one language which do not have any that answer them in another" which means that "those of one country, by their customs and manner of life, have found occasion to make several complex ideas, and given names to them, which others never collected into specific ideas" (Locke 1955: 54–56 [1690] in Wierzbicka 1997). According to Anna Wierzbicka these 'complex ideas' embedded in words are the origin of *interculturalism* via language. For another Polish linguist, Jan Miodek (2016), language is a kind of a mirror in which great political or social emotions are reflected.[3]

There are two important kinds of cultural keys words (Fr. mots chargés culturellement). Firstly, there are untranslatable words like Russian гулаг or Ukrainian майдан. Paradoxically, these words are less difficult to be taught since they represent unique phenomena specific to the cultures they belong to. Such words are integrated into other linguistic systems without any attempt at translation. As a consequence, there is no risk of a linguistic misunderstanding

---

[2] „Kompetencja interkulturowa to zespół umiejętności, które kształtują się, gdy kultura obca 'nakłada' się na kulturę rodzimą uczącego się języka obcego i wpływa na zmianę jego postawy."

[3] „Nagromadzenie wielkich politycznych, społecznych emocji znajduje swoje ujście w języku."

caused by designating a social or cultural phenomenon by a foreign word whose meaning does not correspond to that in the original language.

The second group of lexis includes terms that have their equivalents in a foreign language, easily accessible in bilingual dictionaries. Thus the Polish words *ojczyzna* (Eng. homeland), *przyjaciel* (Eng. friend) or *wolność* (Eng. freedom, liberty) have their equivalents for example in French, English or Russian. However, according to Wierzbicka's study, these equivalents do not have the same meaning. Thus, the American word *freedom* is not equivalent to the Russian word *свобода* or the word *родина* which has a different meaning for Russians than it does for Americans or Poles (Wierzbicka 1997). The intuitive equivalence is illusory and it can only make teaching more difficult instead of facilitating it.

3. The *abstract* and *concrete* in foreign language teaching

As mentioned above, the key words of a culture are often abstract. When we try, however, to teach abstract terms, we 'lose' their true meaning by moving to a certain level of generalization, by moving away from the concrete social context in which they appeared. Looking for a deep meaning of words, foreign language learners cannot be satisfied with general, approximate understanding; they should restore this concrete meaning which reveals the true content of thought (Zarate 1997 in Forestal 1998). Gisela Baumgratz-Gangl (1993) points to a very significant difference between acquiring a mother tongue and a foreign language. While children acquire their first language, they proceed from the *concrete* to the *generalization*. Foreign language learning is characterized by the reverse process. To understand a foreign word that represents a complex and abstract idea, very specific realities that founded this word should be identified and comprehended. One of such words is *heritage* whose French and Polish equivalents – *patrimoine* and *dziedzictwo* – will be analyzed in this paper.

4. Corpora and methodological approach

The study is founded on the following corpora: the Corpora Collection of Leipzig University (LC), the Polish National Corpus (NKJP) and the French Corpus Frantext. The first one is a set of 252 corpora in different languages. It contains texts from generic web pages, Wikipedia and newspapers available on-line (Goldhahn, Eckart, Quasthoff 2012). Its French part (LC(fr)) contains 1,468,766,604 words, whereas the Polish one (LC(pl)), containing 96,476,260 words, is much smaller.

Moreover, the Polish corpus does not contain any research tools allowing the extraction of all the inflected forms of nouns. This disadvantage can be easily remedied by the manual search of all these forms and by summing up the results. The advantage of the Corpora Collection of Leipzig University is the uniformity of the methodology, research tools and statistical measures applied for all the sub-corpora. Moreover, it features a graphical tool presenting the most frequent co-occurrences, which is an important advantage when used for pedagogical purposes.

Frantext is one of the largest French corpora containing texts from the 19th and especially from the 20th and the 21st centuries, but also medieval texts, texts in Middle French, French pre-classical and classical. 90% of texts are considered to be literary, whereas the remaining 10% are scientific and technical.

The list of sources of NKJP contains not only the classics of Polish literature, but also daily and specialist press, recordings of conversations, leaflets and online texts.

Frantext and the Polish National Corpus are more similar in terms of size. The first one contains 254,000,000 words and the second 240,192,461 words. The principal drawbacks of comparing these corpora are their inequivalent research tools and statistical collocation measures, which can be overcome by some additional calculations, which are necessary for making the comparisons that I will present below.

4.1. Extraction of collocations in a corpus

As I have already noted, an adequate development of lexical competence is crucial for foreign language teaching and the "holy grail of lexical acquisition is the acquisition of meaning" (Manning, Schütze 1999). However, to find the meaning of words, especially of cultural key words, is a big challenge for language teaching in the intercultural approach. A. Koselak notes that "it seems logical that the most frequent lexemes designate the most salient realities" (2003). J.-L. Dessalles clarifies that "the secret of extracting the meaning of words lies in their co-occurrence"[4] (2019). Consequently, we will analyze frequent adjectival collocations of the words *patrimoine* and *dziedzictwo* to shed light on their meaning and to reveal what salient cultural realities they hide, which could be instructive for foreign language learners and would give them an insight into French culture.

The term 'collocation of a word' is defined as "statements of the habitual or customary places of that word" (Manning, Schütze 1999). To identify collocations in a corpus, one must rely on appropriate measures of the 'attraction between words' that let predict the neighbourhood of a word (Evert 2007, Bouma

---

[4] «Le secret de l'extraction du sens des mots réside dans leur cooccurrence» (Dessalles 2019).

2009). The choice of measures depends on many factors and can sometimes lead to contradictory results. It becomes even more complicated when we want to compare the results of analyses of several corpora since different corpora provide different measures of collocations. These aspects need to be explained before learners start exploring corpora by themselves and drawing conclusions, especially in the contrastive perspective.

The simplest and most intuitive measure is the frequency (Manning, Schütze 1999) since it assumes that "if two words occur together a lot, then that is evidence that they have a special function" (ibidem). This measure is available in Frantext and in NKJP, as they provide the information about the frequency of co-occurring words, unlike the LC corpora.

The disadvantage of the frequency criterion is that it returns grammatical words as the most frequent, which can be partly remedied by the use of a filter of parts of speech (ibidem). The other issue is that "if the two constitute words of a frequent bigram like *new companies* are frequently occurring words (as new and companies are), then we expect the two words to co-occur a lot just by chance, even if they do not form a collocation" (ibidem). However, they can also form a collocation and that is why frequent n-grams should be analyzed. For less frequent words, more sophisticated methods of the collocation extraction are used. The measures such as MI, Chi^2 or log-likelihood are very widespread as convenient measures of the force of the 'attraction between words'. However, very frequently used words do not always form strong collocations in terms of very high values of the above measures. I therefore chose to compare the results of collocation analysis based on different collocation measures, including frequency. Moreover, I will calculate some measures unavailable in the studied corpora to make my contrastive analysis more reliable and more complete.

4.2. *Dziedzictwo* vs. *patrimoine* in LC, NKJP and Frantext

In this subchapter I will compare the collocations extracted from 4 corpora LC(fr), LC(pl), NKJP and Frantext. Firstly, the collocations will be based on the collocation measures offered by these corpora. Thus, in LC and Frantext, the collocation measure is log-likelihood (LLR), whereas in NKJP it is Chi^2.

The first step is the analysis of the most frequent co-occurrences of the Polish word *dziedzictwo* (nominative case, singular) in LC(pl) and of the French word *patrimoine* (singular), presented in the two figures below. The width of the lines in the graphs represents the strength of the significance of the co-occurrence of the words in question.

Figure 1: Co-occurrences of *dziedzictwo*.
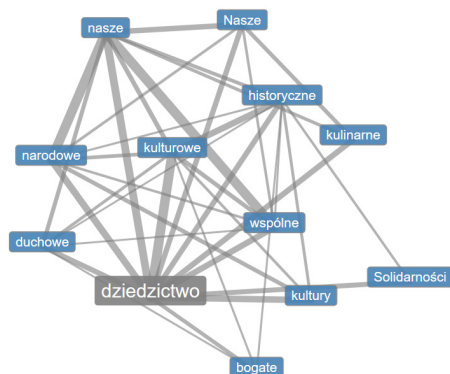
Calculated parameters:
frequency class: (all cases, plural and singular included)   – 11
number of occurrences (all cases, plural and singular included)   – 1839
frequency per million words (fmw)   – 19

The total number of the occurrences of *dziedzictwo* is 1839. It was calculated manually for all cases, for plural and singular, as there is no such a tool in the corpus. To compare the frequencies of the analyzed words, the frequency per million words (fmw) was calculated – 19. The parameter offered by the corpus, which also enables frequency comparison, is the frequency class. It indicates words of similar frequency and it rarely changes between corpora of the same language. Most frequent words have lower classes, whereas very rare words can have "a frequency class of 20 or higher in large corpora", as we can read in the description of the LC corpus. The frequency class of *dziedzictwo* is 11. The most frequent adjectival co-occurences are *kulturowe* (En.cultural), *narodowe* (En. national), *kulinarne* (En. culinary) and *duchowe* (En. spiritual).

Figure 2: Co-occurrences of *patrimoine*.
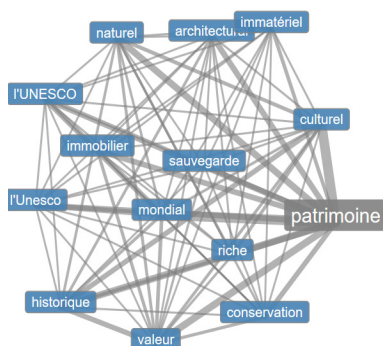
Calculated parameters:

frequency class: (singular and plural included)	– 10
number of occurrences (singular and plural included)	– 76700
frequency per million words (fmw)	– 52

The total number of the occurrences of *patrimoine* is 76700. It was calculated manually for plural and singular. The frequency per million words (fmw) is 52. The frequency class of *patrimoine* is 10. The most frequent adjectival co-occurences are *mondial* (En. global), *culturel* (En. cultural), *naturel* (En. natural) and *architectural* (En. architectural).

According to LC(fr) and LC(pl), *patrimoine* is more frequent than *dziedzictwo* in terms of the frequency per million words and the frequency class. What is more, the common shared adjectival co-occurrences for *patrimoine* and *dziedzictwo* are *cultural* (Pl. kulturowe, Fr. culturel), *historical* (Pl. historyczne, Fr. historique) and *rich* (Pl. bogate, Fr. riche).

For *dziedzictwo*, I also note two co-occurrences which seem very intuitive for a native speaker of Polish – *national* (Pl. narodowe) and *spiritual* (Pl. duchowe). On the other hand, there is an interesting Polish collocation *culinary heritage* (Pl. dziedzictwo kulinarne) not present in French, which is surprising since it is rather the French tradition which is frequently associated with 'cuisine'.

The next step is to compare the results obtained from these two corpora to those from Frantext and NKJP.

Table 1: *Patrimoine* and *dziedzictwo* in Frantext, NKJP and LC.

Patrimoine	Dziedzictwo
N (number of words) in Frantext: 254,000,000	N (number of words) in NKJP: 240,192,461
f (frequency): 1398	f : 4347
f per million words (fmw(Frantext)): 5,5	f per million words (fmw(NKJP)): 18,1
The strongest adjectival colloca-tions in Frantext (log-likelihood): common (Fr. commun), national (Fr. national), cultural (Fr. culturel), ancient (Fr. ancien), artistic (Fr. artistique)	The strongest adjectival collo-cations in NKJP (Chi^2): cultural (Pl. kulturowe), national (Pl. narodowe), culinary (Pl. kulinarne) global (Pl. światowy), natural, biological (Pl. przyrodnicze)
Leipzig corpora (LC(FR))	**Leipzig corpora (LC(PL))**

Patrimoine	Dziedzictwo
N : 1,468,766,604	N : 96,476,260
f : 76700	f : 1839
f per million words (fmw(fr)): 52	f per million words (fmw(pl)): 19
The strongest adjectival colloca-tions in LC(fr) (log-likelihood): global (Pl. światowy), cultural (Fr. culturel), architectural (Fr. architectural), natural (Fr. naturel), historical (Fr. historique)	The strongest adjectival colloca-tions in LC(pl)(log-likelihood): cultural (Pl. kulturowe), national (Pl. narodowe), culinary (Pl. kulinarne) spiritual (Pl. duchowe) common (Pl. wspólne)

(The co-occurrences referring to *gene heritage* (Fr. patrimoine génétique, Pl. dziedzictwo genetyczne) and real estate (Fr. patrimoine immobilier, Pl. nieruchomości) were not calculated among both the Polish and French co-occurrences, as my research is focused only on the cultural component of the meaning of the words in question.)

Firstly, according to the above data, it is impossible to decide which word – *patrimoine* or *dziedzictwo* – is more frequently used since the results in LC(pl) and LC(fr) differ from those extracted in Frantext and NKJP. Thus, the French word *patrimoine* is more frequent in LC(fr) than the Polish word *dziedzictwo* in LC(pl), and its fmw(fr) (52) is higher than fmw(pl) of *dziedzictwo* (19). This is confirmed by the calculation of the frequency class (10) of *patrimoine,* which is lower than the frequency class of *dziedzictwo* (11). However, the comparison of the fmw in Frantext and NKJP gives opposite results.

Secondly, the common adjectival collocation for *patrimoine* and *dziedzictwo* in all analyzed corpora is *cultural* (Pl. kulturowe, Fr. culturel). The adjective *national* (Pl. narodowe, Fr. national) is present in 3 of the above cases. The collocates such as *global* (Fr. mondial), *common* (Fr. commun), and even *natural* (Fr. naturel, Pl. przyrodniczy (possible translation)) are also present among the strongest collocations of *dziedzictwo* in NKJP or LC(pl), whereas *architectural* (Fr. architectural) and *artistic* (Fr. artistique) are strong collocates only for *patrimoine*. Similarly, *culinary* (Pl. kulinarny) is a strong collocate only for *dziedzictwo*. It should also be noted that the results from NKJP and LC(pl) are quite convergent as far as the three strongest collocations are concerned – *cultural* (Pl. kulturowe), *national* (Pl. narodowe) and *culinary* (Pl. kulinarne).

4.3. Comparison of the results of different collocation measures

As it has been already mentioned, comparing the collocations obtained by different collocation tests should be avoided, as we risk taking into consideration different linguistic phenomena. On the one hand, some collocation tests are more "conservative" as they correlate better at a high frequency (as log-likelihood (LLR)). As Moore (2004) says, "its use remains controversial on the grounds that it may be unreliable when applied to rare events". To use uniquely LLR can be an inconvenience since rare words comprise a large and important portion of terminology, as Zipf's law suggests (Thanopoulos, Fakotakis and Fokkinakis, 2002). On the other hand, other tests are more 'sensitive' to lower frequencies of analyzed pairs (as very frequently used Mutual Information (MI)) (Rudolf, 2004). That is why I decided to compare collocations based on the same collocation measures, which can be easily calculated, given the frequency of analyzed words. These measures are MI, LogDice, MI^2, MI^3, where MI is the most 'sensitive', and MI ^3 the most 'conservative' as also confirmed by the calculation of the correlation coefficient for different collocation measures and the frequencies of the adjectival collocates (in tables 3 and 4) presented in the table below.

Table 2: Comparison of the correlation coefficient of different collocation measures.

Correlation coefficient (Excel) for *patrimoine* in Frantext		Correlation Coefficient (Excel) for *dziedzictwo* in NKJP	
CorCoef(f|MI^3)	0,679148	CorCoef(MI^3)	0,6819
CorCoef (f|MI^2)	0,519813	CorCoef(MI^2)	0,58107
CorCoef (f|LogDice)	0,318109	CorCoef(LogDice)	0,54108
CorCoef (f|MI)	0,250743	CorCoef(MI)	0,32343

In the table below, I present the most frequent adjectives (frequency greater or equal to 5) co-occurring with *patrimoine* in Frantext, sorted by frequency.

Table 3: The most frequent adjectives co-occurring with *patrimoine* in Frantext.

Adjective	f(adj|patrimoine)	Collocation measures			
		MI	MI^2	MI^3	LogDice
commun (common)	**42**	7,37	**12,76**	**18,15**	4,86
national	25	7,49	12,13	16,78	4,94
ancien (ancien)	22	5,89	10,35	14,81	3,39
culturel	**19**	**10,12**	**14,37**	**18,62**	**7,11**

Adjective	f(adj| patrimoine)	Collocation measures			
		MI	MI^2	MI^3	LogDice
propre (own)	<u>14</u>	5,03	8,84	12,64	2,53
artistique	**13**	**<u>9,68</u>**	**<u>13,38</u>**	**<u>17,08</u>**	**<u>6,64</u>**
spirituel	9	7,18	10,35	13,52	4,54
intellectuel	8	7,2	10,2	13,2	4,54
français	8	4,64	7,64	10,64	2,14
petit (small)	8	3,86	6,86	9,86	1,37
touristique	**7**	**<u>10,75</u>**	**<u>13,56</u>**	**<u>16,37</u>**	**<u>6,75</u>**
moral	7	5,91	8,71	11,52	3,34
seul (only)	7	3,67	6,48	9,28	1,18
forestier (forest)	**6**	**<u>9,72</u>**	**<u>12,3</u>**	**14,89**	**<u>6,19</u>**
familial (family)	**6**	**<u>7,81</u>**	**10,39**	**12,98**	**<u>4,97</u>**
riche	6	5,27	7,85	10,44	2,73
beau (beautiful)	6	3,45	6,03	8,62	0,96
privé (private)	5	6,55	8,87	11,19	3,88
naturel	5	4,34	6,66	8,98	1,82
autre (other)	5	3,18	5,51	7,83	0,69

Although all the above results show that the extraction of collocations depends on collocation measures, it is possible to note some regularities. For MI, LogDice and MI^2, the strongest collocates are *touristique*, *culturel*, *forestier*, *artistique*. Among these words, three adjectives *touristique, culturel, artistique* are indicated as strong collocates also by MI^3. It should be noted that these words do not have the highest frequency. Let us recall that the three strongest collocations extracted by LLR from Frantext are not the same – *common* (Fr. commun), *national* (Fr. national) and *cultural* (Fr. culturel).

In the table below, I present the adjectives (frequency greater or equal 5) co-occurring most frequently with *dziedzictwo* in NKJP sorted by frequency.

Table 4: The most frequent adjectives co-occurring with *dziedzictwo* in NKJP.

Przymiotnik (Eng. Adj)	f (adj| dziedzictwo)	MI	LogDice	MI ^2	MI ^3
narodowe (national)	**1011**	**10,01**	**9,15**	**20,00**	**29,98**
kulturowe, kulturalne (cultural)	**739**	**11,17**	**10,10**	**20,70**	**30,23**
światowe (global)	**248**	8,75	7,81	**16,70**	**24,66**
polskie (Polish)	**86**	3,84	3,07	10,27	16,69
europejski (European)	**69**	5,74	4,90	11,85	17,96
wspólny (common)	63	6,17	5,29	12,15	18,13
przyrodnicze (natural, biological)	58	9,78	**7,89**	15,63	21,49
historyczne (historical)	50	7,21	6,16	12,86	18,50
kulinarne (culinary)	46	**10,39**	**7,92**	15,91	21,44
chrześcijańskie (Christian)	43	**11,35**	**8,07**	**16,78**	**22,21**
duchowe (spiritual)	40	7,58	6,37	12,90	18,23
wielkie (big)	36	3,74	2,95	8,91	14,08
całe (whole)	27	2,63	1,85	7,39	12,14
własne (own)	24	3,45	2,65	8,04	12,62
bogate (rich)	21	6,14	5,05	10,53	14,92
żydowskie (Jewish)	18	6,54	5,29	10,71	14,88
naturalne (natural)	15	5,09	4,10	9,00	12,90
komunistyczne (communist)	13	6,48	5,11	10,18	13,89
archeologiczne (archeological)	12	8,45	5,98	12,04	15,62
regionalne (regional)	11	4,97	3,92	8,43	11,89
przemysłowe (industrial)	10	5,88	4,58	9,20	12,53
trwałe (lasting, enduring)	10	4,62	3,61	7,94	11,26
poprzednie (previous)	10	3,90	2,98	7,22	10,54
greckie (Greek)	9	6,59	4,96	9,76	12,93
materialne (material)	9	5,50	4,27	8,67	11,84

Przymiotnik (Eng. Adj)	f (adj| dziedzictwo)	MI	LogDice	MI ^2	MI ^3
chopinowskie (Chopin)	8	9,57	5,73	12,57	15,57
wspaniałe (great)	8	4,41	3,38	7,41	10,41
dawne (ancient)	8	2,91	2,05	5,91	8,91
jedno (one)	8	0,17	-0,60	3,17	6,17
romantyczne (romantic)	7	6,52	4,74	9,32	12,13
rzymskie (Roman)	7	6,12	4,53	8,92	11,73
smutne (sad)	7	5,09	3,87	7,90	10,71
lokalne (local)	7	3,77	2,81	6,58	9,38
rodzinne (family)	7	3,72	2,77	6,53	9,34
śląskie (Silesian)	7	3,36	2,45	6,17	8,97
dobre (good)	7	1,37	0,58	4,18	6,99
rodowy (ancestral)	6	**<u>18,29</u>**	5,50	**<u>20,87</u>**	**<u>23,46</u>**
wielowiekowe (centuries-old)	6	9,09	5,31	11,67	14,26
kolonialne (colonial)	6	8,23	5,17	10,81	13,39
architektoniczne (architectural)	6	7,06	4,85	9,64	12,23
realne (real)	6	4,64	3,48	7,22	9,81
naukowe (scientific)	6	4,03	3,00	6,62	9,20
niemieckie (German)	6	3,06	2,15	5,64	8,23
ogromne (huge)	6	3,03	2,13	5,62	8,20
polityczne (political)	6	2,40	1,55	4,98	7,57
śródziemnomorskie (Mediterranean)	5	8,34	4,98	10,66	12,98
antyczne (ancient)	5	7,55	4,81	9,87	12,19
stalinowskie (Stalinist)	5	7,30	4,75	9,62	11,94
kłopotliwe (embarrassing)	5	6,83	4,59	9,15	11,47
cywilizacyjne (civilizational)	5	6,71	4,55	9,03	11,35
pruskie (Prussian)	5	6,52	4,48	8,84	11,16
ukraińskie (Ukrainian)	5	5,74	4,11	8,06	10,38

Przymiotnik (Eng. Adj)	f (adj| dziedzictwo)	MI	LogDice	MI ^2	MI ^3
małopolskie (Lesser Poland)	5	5,69	4,08	8,01	10,34
intelektualne (intellectual)	5	5,35	3,88	7,67	9,99
wieczne (eternal)	5	5,16	3,76	7,48	9,80
klasyczne (classical)	5	4,72	3,46	7,04	9,36
literackie (literary)	5	4,57	3,36	6,89	9,21
religijne (religious)	5	4,37	3,21	6,69	9,01
minione (passed)	5	4,27	3,13	6,59	8,91
wszelkie (all kinds)	5	2,60	1,72	4,93	7,25
złe (bad)	5	1,94	1,11	4,27	6,59
trudne (difficult)	5	1,54	0,72	3,86	6,18
częste (frequent)	5	0,99	0,19	3,31	5,63

The first observation is that there are more Polish collocations (63) than French ones (20).

Secondly, the discussion of the analysis of adjectival collocations for the Polish word *dziedzictwo* reveals the same doubts as described for *patrimoine*. Among the five strongest collocations, there are three adjectives – *national* (Pl. narodowe), *cultural* (Pl. kulturowe, kulturalne), and *Christian* (Pl. chrześcijańskie). The adjectives *national* and *cultural* are also the most frequent. This analysis revealed that the adjective *ancestral* (Pl. rodowe), which is indicated as a strong collocate by three measures (MI, MI^2 and MI^3), does not form a frequent collocation (6 occurrences). Afterwards, *global* (Pl. światowe), which forms a very frequent collocation (258 occurrences), is indicated by two measures, similarly to the adjective *culinary*, forming a less frequent collocation (46 occurrences). It should be noted that there is some convergence among the Polish collocations extracted from NKJP by my calculations and Chi^2 – the three strongest collocations indicated by Chi^2 in NKJP are *cultural* (Pl. kulturowe), *national* (Pl. narodowe) and *culinary* (Pl. kulinarne).

Having compared the results obtained from NKJP and Frantext, the word *cultural* forms a strong common collocation. In French, other collocates – *artistic* and *touristic* diverge with other collocates of *dziedzictwo* – *narodowe* and *chrześcijańskie*.

Afterwards, I calculated the correlation coefficient based on the frequency of the adjectival collocations of *patrimoine* and *dziedzictwo,* respectively, in Frantext

and NKJP. In this case I also took into consideration frequencies less than 5, for those adjectives that are frequent for one of the two words. If some adjectives appeared for example only for *patrimoine* (with the frequency greater or equal to 5) but they did not appear for *dziedzictwo*, its frequency for *dziedzictwo* was considered to be 0. As a result I obtained the correlation coefficient:

$$K = 0,4$$

which indicates a weak level of correlation.

4.4. Categorization of adjectival co-occurrences

In this section, I will analyze the most frequent adjectives co-occurring with *patrimoine* and *dziedzictwo*, based on Frantext and NKJP. I distributed all the adjectives occurring five or more times in eight categories: cultural, national, collective, activities, spiritual, qualitative, time and other. I obtained the results presented in the figure below.

Figure 3: Categorization of the adjectival collocations of *patrimoine* and *dziedzictwo* in Frantext and NKJP.
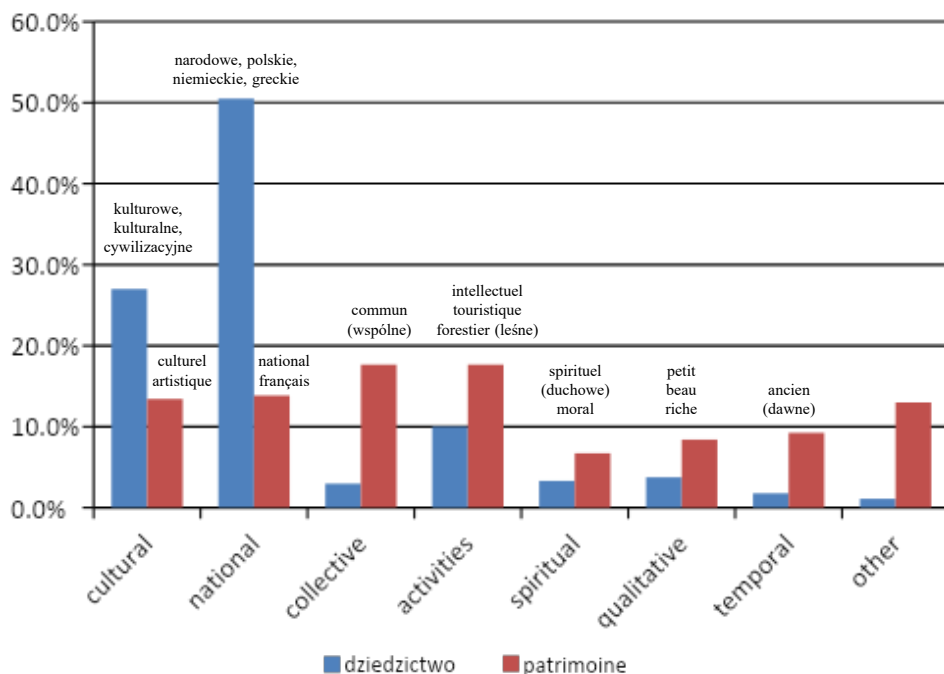
(The data in figure 3 is normalized to the total number of all the adjectival co-occurrences in each language.)

The first observation is that the distribution of the above categories for *patrimoine* is more uniform, whereas for *dziedzictwo*, there are two categories dominating (cultural and national). Furthermore, the figure shows that *patrimoine* occurs more frequently than *dziedzictwo* with adjectives from the following categories: collective, activities, spiritual, qualitative, temporal and other. These adjectives designate 'more concrete meaning', whereas *dziedzictwo* is dominating in "more abstract" categories such as cultural or national. This 'less abstract' character of the French *patrimoine* can also be better understood if we mention some adjectives that co-occur seldom with *patrimoine* in LC(fr) and that hardly ever occur in Polish: *patrimoine minier* (Pl. dziedzictwo kopalniane), *audiovisuel* (Pl. dziedzictwo audiowizualne), *communal* (Pl. komunalne), *cinématographique* (Pl. kinematograficzne). Although they are not strong collocates in French, they are used in communication and consequently complete the meaning of the French *patrimoine*.

5. Conclusion

My research permits to draw some methodological and intercultural conclusions that can facilitate the integration of corpus analysis in the intercultural approach in foreign langue teaching.

First of all, some collocation measures return function words as the most frequent. Consequently, there is the necessity to extend the analysis to less frequent co-occurring words. Then, some corpora offer tools which do not include all of the inflected forms of words, for instance the frequency class in LC. Next, to compare corpus data form different corpora, an appropriate solution is to use the same collocation measures, hence the huge advantage of multilingual corpora such as LC. The presence of graphic tools visualizing the relations between words is also very advantageous for didactic applications.

As for corpora like NKJP and Frantext, which do not use the same collocation measures, users should be very cautious when comparing the strength of collocations. It is more reliable to use several different collocation measures, for two reasons. Firstly, if a few collocation measures allow the extraction of the same collocations, they can be considered as strong with a high degree of certainty. In our case it was ***cultural heritage*** – a very frequent and strong collocation for both *patrimoine* and *dziedzictwo* indicated by all the collocation measures. Then, the other strong Polish collocations are ***national heritage*** and ***Christian heritage***, although the latter was not detected by all the collocation measures, whereas in French these are ***touristic*** and ***artistic heritage***. Secondly, such an analysis can return interesting collocations, despite not being very frequent. In the case of *dziedzictwo*, it was *ancestral heritage* (Pl. dziedzictwo rodowe).

The collocate *Christian* (Pl. chrześcijański), which was not indicated as a very strong collocate by Chi^2 in NKJP but returned by my additional analysis, seems very significant for the Polish *dziedzictwo*. The French *patrimoine* seems strongly associated with artistic and touristic domains, which is not observed for the Polish collocations of *dziedzictwo*. As art and tourism topics are very often present in students' books for French language teaching, it is important to attract students' attention to the importance of these collocations in French and the difference in meaning between Polish and French *heritage* in general.

As a result of my analysis, it appeared, surprisingly, that *culinary heritage* (Pl. dziedzictwo kulinarne) is quite a strong collocation that does not occur among the strong French collocations.

The categorization of the analyzed collocates showed that the Polish *dziedzictwo* is more often used in more 'abstract contexts', whereas the French *patrimoine* occurs in more 'concrete ones'. It seems that in French, nearly everything important for a group of people can become *patrimoine*, whereas in Polish, *dziedzictwo* is associated with domains of greater importance for bigger groups of people or the whole nation.

As for applying a collocation analysis in foreign language teaching, two approaches can be considered. In the first one, the students search for collocations by themselves using collocation measures offered by the analyzed corpora. This approach needs the preliminary introduction of the notion of collocation as well as basic knowledge concerning collocation measures, presented in my first conclusion. In the second approach, the results of my analysis can be directly incorporated into the curriculum. Both approaches should be accompanied by a profound reflection focused on the relation between the language and the culture, on cultural key words and on the place of heritage in the contrastive perspective. The last stage would be the most crucial part of both approaches, as it contributes directly to the development of the intercultural competence and the formation of the intercultural personality of learners.

Bibliography

Abramowicz Maciej (1993): Patrie – „ojczyzna"? In: Bartmiński Jerzy (eds): *Pojęcie ojczyzny we współczesnych językach europejskich*. Lublin, 241–248.

Bartmiński Jerzy (1993): Polskie rozumienie ojczyzny i jego warianty. In: Bartmiński Jerzy (eds): *Pojęcie ojczyzny we współczesnych językach europejskich*. Lublin, 23–48.

Baumgratz-Gangl Gisela (1993): *Compétence transculturelle et échanges éducatifs*. Paris.

Bouma Gerlof (2009): *Normalized (Pointwise) Mutual Information in Collocation Extraction*. Potsdam: https://pdfs.semanticscholar.org/1521/8d9c029cbb903ae7c729b2c644c24994c201.pdf (13.03.2020).

Council of Europe (2001): *Common European Framework of Reference for Languages: learning, teaching, assessment (CEFR)*. Strasbourg.

Dessalles Jean-Louis (2019): *Des intelligences très artificielles*. Paris.

Dryjańska Agnieszka (2016): Słowa-wartości: interkulturowość ukryta w językach. *Neofilolog*, *47*, *1*, 37–51.

Evert Stefan (2007): *Corpora and collocations*. Osnabrück: http://www.stefan-evert.de/PUB/Evert2007HSK_extended_manuscript.pdf (13.03.2020).

Forestal Christine (1998): L'éthique au cœur de la didactique des langues. *ÉLA. Études de linguistique appliquée*, *109*, 49–61.

Francesco Attruia (2010): De l'approche lexiculturelle des dictionnaires monolingues: le nouveau Petit Robert 2010, version électronique. *ÉLA. Études de linguistique appliquée*, *157*, 9–22: https://www.cairn.info/revue-ela-2010-1-page-9.htm (13.03.2020).

Galisson Robert (1998): Cultures et lexicultures. Pour une approche dictionnairique de la culture partagée. *Cahiers d'Études Hispaniques Médiévales*, 325–341: https://www.persee.fr/doc/cehm_0180-9997_1988_sup_7_1_2133 (13.03.2020).

Galisson Robert (2005): La pragmatique lexiculturelle pour accéder autrement à une autre culture, par un autre lexique. *Mélanges CRAPEL, 25*, p. 47–73: http://www.atilf.fr/spip.php?article3471 (13.03.2020).

Gallon Fabienn, Himber Céline (2016): *En Action 2*. Vanves.

Goldhahn Dirk, Eckart Thomas, Quasthoff Uwe (2012): Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation* (LREC12).

Kłoskowska Agnieszka (1993): Kraj, do którego się wraca. Czym jest ojczyzna dla lubelskich studentów. In: Bartmiński Jerzy (eds): *Pojęcie ojczyzny we współczesnych językach europejskich*. Lublin, 49–56.

Koselak Arkadiusz (2003): La sémantique naturelle d'Anna Wierzbicka et les enjeux interculturels. *Questions de communication, 4*, 83–95: https://doi.org/10.4000/questionsdecommunication.4611 (13.03.2020).

Manning, Christopher D., Schütze Hinrich (1999): *Foundations of Statistical Natural Language Processing*. London: https://www.cs.vassar.edu/~cs366/docs/Manning_Schuetze_StatisticalNLP.pdf (13.03.2020).

Miodek Jan (2016): Proszę przy mnie…: https://plus.dziennikzachodni.pl/miodek-prosze-przy-mnie-nie-uzywac-tego-slowa-nie-cierpie-go/ar/10736192 (13.03.2020).

Moore Robert C. (2004): On Log-Likelihood-Ratios and the Significance of Rare Events: https://www.aclweb.org/anthology/W04-3243.pdf (13.03.2020).

Rudolf Michał (2004): *Metody automatycznej analizy korpusów tekstowych*. Warszawa.

Skibińska Elżbieta (2003): Pojęcie stereotypu w badaniach francuskich. Przegląd najnowszych tendencji. *Etnolingwistyka. Problemy Języka i Kultury*, *15*, 63–79: http://dlibra.umcs.lublin.pl/dlibra/doccontent?id=1486&dirids=1 (13.03.2020).

Thanopoulos Aristomeni, Fakotakis Nikos, Kokkinakis George (2002): Comparative Evaluation of Collocation Extraction Metrics: http://www.lrec-conf.org/proceedings/lrec2002/pdf/128.pdf (13.03.2020).

Wierzbicka Anna (1997): *Understanding Cultures Through Their Key Words, English, Russian, Polish, German, and Japanese*. New York, Oxford: https://archive.org/stream/UnderstandingCulturesThroughTheirKeyWords/Understanding%20Cultures%20Through%20Their%20Key%20Words_djvu.txt (13.03.2020).