

# Annales Universitatis Paedagogicae Cracoviensis

Studia Linguistica 17 (2022)

ISSN 2083-1765

DOI 10.24917/20831765.17.2

**Vitalija Kazlauskienė**

ORCID 0000-0003-0505-7770

Université de Vilnius, Vilnius, Lituanie

**Agnieszka Dryjańska**

ORCID 0000-0003-1649-8408

Université de Varsovie, Varsovie, Pologne

## Le sens de *fête* en polonais, en lituanien, en français et sa (non)coïncidence collocationnelle

**Les mots clés :** analyse de corpus, sémantique distributionnelle, collocation, fête, FLE

**Słowa kluczowe:** analiza korpusowa, semantyka dystrybucyjna, kolokacja, święto, FLE

**Keywords:** corpus analysis, distributional semantics, collocation, holiday, French Language Teaching

### Introduction

La fête suscite l'intérêt de la didactique des langues en stimulant des réflexions linguistiques et interculturelles. Dans les méthodes de français<sup>1</sup>, la notion de 'fête' apparaît, certes, mais elle s'appuie souvent sur des exemples de *fête de la musique*, *des voisins* ou *du village* dont l'importance pour les Français n'est claire ni pour les apprenants lituaniens ni polonais, vu que dans ces deux pays d'autres festivités jouent un rôle plus important. De surcroît, l'épidémie de Covid-19 a révélé un grand besoin social et psychologique de fête, ce dont témoignent des néologismes comme *coronapéro*, *apérue*, *Zoombar* en France et de nombreuses fêtes organisées dès la levée des restrictions sanitaires dans différents pays, entre autres en France, en Pologne et en Lituanie.

L'intérêt pour la fête dans notre recherche est également lié au passé commun de la Lituanie et de la Pologne. Il semblait probable que certaines caractéristiques culturelles aient des points de convergence reflétés dans nos deux langues de manière similaire. Cette attente était aussi renforcée par une correspondance étymologique entre *święto* et *šventė*, les équivalents respectifs polonais et lituanien de *fête*. C'est à ce titre que l'étude des collocations de mots chargés culturellement comme *fête* pourrait être interprétée en tant qu'indice culturel. L'aperçu linguistique de ce mot dans les trois langues s'annonce prometteur dans l'approche interculturelle en didactique du FLE afin de dépasser les contextes *grosso modo* monoculturels en Pologne et en Lituanie.

---

<sup>1</sup> Par exemple dans Alter Ego +, Édito, Latitudes, Sésame, Tendances, Saison.

## Objectif de la recherche

L'objectif de cette recherche est d'analyser les images linguistiques du mot *fête* et de ses équivalents en lituanien et en polonais en se basant sur des corpus de texte en trois langues et de traiter leur (non)coïncidence collocationnelle. Notre étude s'inscrit dans la sémantique distributionnelle, fondée sur une analyse de corpus, à des finalités didactiques.

Pour atteindre l'objectif, conformément à une méthodologie de Manning et Schütze (1999) (« one can categorize a word by the words that occur around it. »), nous allons proposer une méthode de catégorisation des collocations de *fête* relevées dans plusieurs corpus français, polonais et lituanien afin de vérifier dans quelle mesure elles reflètent les différences et ressemblances langagières et culturelles dans les trois pays.

Souvent le mot, dont la signification semble transparente pour les apprenants, dans un autre contexte, langagier ou social, représente une signification inattendue et ce n'est que dans l'ensemble des mots qu'il est possible de la déchiffrer correctement ; parfois d'éliminer la polysémie (Kazlauskienė 2022). La question de la fréquence amène à caractériser les cas particuliers de cooccurrences lexicales en tant que collocations (Ancitil, Tremblay 2016). Les collocations posent souvent problème aux apprenants sur le plan réceptif, mais davantage encore sur le plan productif. Sur le premier plan, le sens global n'est souvent pas prédictible à partir des constituants ou à partir de la langue maternelle. Sur le plan productif, car il n'y a pas de contexte qui permette d'inférer le sens (Binon, Verlinde 2003 : 32).

Afin de bien expliquer la démarche qui a permis d'atteindre l'objectif de notre analyse collocationnelle, nous allons tout d'abord esquisser son cadre théorique en éclairant les nuances terminologiques et conceptuelles liées au terme de *collocation*. Ensuite, après avoir détaillé les aspects méthodologiques, nous présenterons toutes les étapes de l'étude : analyse étymologique, lexicographique et celle de corpus.

## Collocations – cadre théorique

La collocation est un phénomène largement discuté qui est pourtant loin de faire l'unanimité vu ses « contours flous » (Tutin, Grossmann 2002) générant toute sorte de controverses et malentendus (Evert 2007).

On distingue *grosso modo* deux approches<sup>2</sup> : phraséologique (Benigno, Grossmann, Kraif 2015), plus étroite (Tutin, Grossmann 2002), fondée sur les conceptions psychologisantes de Bally (cité par Śliwa 2011), ensuite développées par Mel'čuk (1998) et Hausmann (2007) et la deuxième approche – formalisante (cité par Śliwa 2011) plus large (Tutin, Grossmann 2002), appelée empirique (Evert 2007) ou statistique (Benigno, Grossmann, Kraif, 2015) introduite par Firth (Firth et al. 1962 (1957)), reprise par les fonctionnalistes anglais Halliday (1985), Sinclair (1991) et puis Stubbs (1996), McEnry et Wilson (2001), favorisant la caractéristique quantitative de la collocation et, en conséquence, souvent employée en TAL. Chaque approche

---

<sup>2</sup> Notons la présence de la troisième approche – approche cognitive – représentée, entre autres, par D. Śliwa (2011) qui est devenu une source d'inspiration pour la didactique (Benigno, Grossmann et Kraif 2015).

propose des définitions de la collocation et des terminologies différentes bien qu'il y ait un « overlap between the phraseological notion of collocations and the more general empirical notion » (Evert 2007). En plus des définitions phraséologique et empirique des collocations, Evert remarque que, dans la linguistique computationnelle, le mot *collocation* est souvent utilisé comme *a generic term* désignant toutes les combinaisons lexicalisées (ang. lexicalised word combinations). La collocation « empirique » (Evert 2007) peut être définie simplement comme « des mots qui tendent à apparaître ensemble » (Tutin, Grossmann 2002). On met aussi en valeur le facteur de fréquence en parlant de cooccurrences récurrentes (Evert 2007).

Dans la perspective phraséologique, la notion de collocation se présente souvent comme une cooccurrence lexicale binaire (Tutin 2013 : 49) semi-figée ou « semi-idiomatique » (Tremblay 2014 : 74) : une base et un collocatif, où une base conserve son propre sens, tandis qu'un collocatif complète son sens à la manière particulière, contrainte de manière faible (*grande nouvelle*) ou plus forte (*rire à gorge déployée*) (Tremblay 2014 : 74). Il est à noter qu'une collocation pourra aussi avoir comme collocatif une collocation (Tutin 2013 : 54), par exemple : *avoir un geste déplacé* (Siepmann 2006), *tomber dans l'oubli*, etc. Ces combinaisons de collocations, assez courantes dans la langue, sont décrites par F. Hausmann (2007) à l'aide du terme de « chaîne collocationnelle ». Ensuite, « il est souvent impossible d'interchanger des collocatifs, même si ceux-ci servent à exprimer une idée semblable » (Tremblay 2014 : 75) ce qui a des conséquences directes pour la didactique des langues.

Pour atteindre le niveau du FLE le plus élevé, ce qui est l'objectif majeur de l'enseignement académique, l'apprenant d'une langue étrangère devrait savoir manipuler correctement les deux types de collocations. Pour la didactique des langues, l'analyse collocationnelle a donc un double intérêt. Tout d'abord, elle fournit des patrons de mots qui sont couramment utilisés dans une langue, et par conséquent doivent être inclus dans l'apprentissage. Ensuite, elle constitue un outil qui permet d'accéder au sens d'un mot conformément à la constatation de Firth (Firth et al. 1962 : 11) « You shall know a word by the company it keeps ». Il en résulte que deux mots qui partagent « le même profil distributionnel [...] sont également proches par le sens » (Fabre 2015), ce qui est utilisé dans la Leipzig Corpora Collection pour relever des synonymes du corpus et a permis de développer le concept des plongements lexicaux (ang. word embeddings) (Mikolov et al. 2013). Ce concept récemment développé commence à susciter l'intérêt des chercheurs en didactique des langues (Rodríguez-Fernández et al. 2016).

### **Caractéristique des collocations**

La collocation est une « unité linguistique qui a plusieurs aspects » (Śliwa 2011) et nombre de linguistes ont essayé de les distinguer de l'expression idiomatique et du nom composé (ibidem). Or, le degré de figement, déjà problématique dans l'approche phraséologique, différenciant les collocations des mots composés, des expressions idiomatiques et des associations libres n'est pas une catégorie facilement applicable dans l'approche quantitative. Dans le cadre purement empirique, quand on extrait automatiquement des « séquences répétées intéressantes » (Silberztein

2015) d'un corpus de texte, on utilise les termes *collocation*, *cooccurrence*, *voisinage* (Frantext), *cooccurrences* et *voisin* (Leipzig Corpora Collection), uniquement *collocation* (Narodowy Korpus Języka Polskiego) et *word sketch* (Sketch Engine) (« word's collocational behaviour » (Thomas 2015)).

## Méthodologie de la recherche

Pour atteindre le but de ce travail, la méthodologie inductive corpus-driven (Tognini-Bonelli 2001) a été choisie. Cette méthodologie nous permettra de découvrir une vraie image linguistique de la fête fondée sur l'analyse de corpus : nous utiliserons des données linguistiques, de plusieurs corpus français, lituanien et polonais pour les catégoriser et analyser dans une perspective contrastive. Selon Teubert (2009), une certaine prudence par rapport à l'analyse *corpus-driven* est recommandée, ce qui se traduit dans notre analyse par une prise en compte d'un savoir anthropologique, culturel et social concernant le phénomène festif dans les trois pays.

Notre démarche se veut semi-automatique : elle est fondée sur des outils d'extraction automatique de collocations et des concordanciers, mais elle inclut un traitement manuel des données. La mesure d'extraction de collocations se base sur la fréquence (fréquence minimale correspond à 5 occurrences). L'analyse empirique des collocations a été basée sur les données des corpus nationaux polonais et lituanien, pour le français sur le corpus de Frantext, et aussi deux groupes de corpus tels que Sketch Engine et Leipzig Corpora Collection. Chaque étape de cette recherche a été élaborée à l'aide des corpus particuliers, de la manière suivante :

A l'étape introductive, l'analyse se base sur le corpus de Leipzig Corpora Collection:

- Leipzig Corpora Collection (corpus lituanien – 4 680 309 phrases 66 635 052 mots);
- Leipzig Corpora Collection (corpus polonais – 96 476 260 mots);
- Leipzig Corpora Collection (corpus français 1 468 766 604 mots).

L'analyse *sensu stricto* prévoit le travail avec des corpus nationaux<sup>3</sup> :

- Frantext – (254,000,000 mots), 90% de textes littéraires, 10% de textes scientifiques et techniques;
- le Corpus national de la langue polonaise (NKJP) – (240 192 461 de mots), textes littéraires, textes de la presse spécialisée, textes d'Internet.
- le Corpus du lituanien contemporain *Dabartinės lietuvių kalbos tekstynas (DLKT)* (140 921 288 mots; sources – textes (non) littéraires, textes de presse, textes administratifs, texte de langue parlée).

Et finalement l'analyse complémentaire se développe à la base de Sketch Engine :

- Sketch Engine lt. – 162.89 ;
- Sketch Engine fr. – 5,752,261,039 ;
- Sketch Engine pl. – 4,253,636,443.

---

<sup>3</sup> NKJP est le corpus national, Frantext n'est pas un corpus national, mais il ressemble au corpus national polonais, DLKT est le corpus du lituanien contemporain.

Il est à noter que les difficultés du traitement des données polonaises et lituaniennes étaient beaucoup plus fréquentes que celles du français, ce qui est évident en prenant en considération des formes fléchies de différents cas grammaticaux de deux langues.

Les occurrences des mots *fête*, *święto* et *šventė* dans les corpus différents sont les suivantes:

NKJP – *święto* ‘fête’, avec toutes formes fléchies, apparaît 26703 fois dans le corpus NKJP.

Frantext – 19243 occurrences de *fête*.

Pour le mot lituanien *šventė* ‘fête’:

DLKT – 3249 (*šventė* ~ *švente* ~ *Šventė* ~ *Švente*)

Sketch Engine – 10367

Leipzig Corpora Collection – 9458

Le nombre des occurrences du mot *šventė* tirées du corpus national lituanien est beaucoup plus restreint que celui de Frantext ou NKJP, en conséquence nous avons analysé parallèlement pour le lituanien les données de Sketch Engine et Leipzig Corpora Collection, en ayant pour but d'équilibrer les données de différents corpus.

### Analyses étymologique et lexicographique de *fête*, *święto* et *šventė*

En analysant l'image linguistique d'un mot, on commence souvent par des définitions lexicographiques (Royle et al. 2009) et son étymologie (Cholewa 2008). Dans notre cas, cette dernière révèle certaines différences significatives dont les conséquences, dans une certaine mesure, seront confirmées par l'analyse de corpus.

Le substantif français *fête* vient du latin *festā* ou *festā dies* signifiant ‘jours de fête’ (Antidote) et est définie comme un « ensemble de réjouissances collectives destinées à commémorer périodiquement un événement » (TLFi). Ce dictionnaire fournit de nombreuses collocations de *fête*, certes, mais il ne permet pas de voir leurs fréquences dans la langue, par conséquent, l'on ne peut pas décider lesquelles sont déjà tombées en désuétude (Royle et al. 2009), donc elles ne devraient pas occuper la première place en FLE.

L'ancien adjectif polonais *święty* ‘saint’ est d'origine protoslave et signifiait *mocny, silny* ‘fort’. L'adjectif concret a pris le sens abstrait désignant *mocny duchem* ‘fort d'esprit’ sous l'influence du latin *sanctus* dans le cadre de la christianisation (Długosz-Kurczabowa 2004). *Święto* désigne « un jour quand on célèbre des festivités religieuses, d'État ou un événement important »<sup>4</sup> (Doroszewski online, traduction).

*Šventė* en lituanien a la même racine, dont le premier sens commun serait « fait d'accomplir un acte sacré » – donc de célébrer la fête. Dans les langues baltes, l'on ne trouve pas d'explication du sens du mot sacré – dans le proto-balte, il est déjà utilisé dans le même sens de « sainteté » (Razauskas 2016 : 15).

<sup>4</sup> pl. « dzień, w którym obchodzi się uroczystości religijne lub państwowe » lub « nadzwyczajne wydarzenie »

## **Analyse de corpus – occurrences de *fête*, *święto* et *šventė* dans Leipzig Corpora Collection, Frantext, NKJP, DLKT et Sketch Engine**

### **Extraction de collocations d'un corpus de texte**

Les corpus offrent différents outils d'extraction de « séquences intéressantes » (Silberztein 2015). Pour les identifier dans un corpus, il faut s'appuyer sur une ou plusieurs mesures statistiques qui quantifient la force d'attraction entre les mots (ang. *attraction between words*) (Evert 2007 ; Bouma 2009).

Ces outils sont fondés sur différents critères parmi lesquels, tout d'abord, la fréquence, critère le plus intuitif, mais non pas le plus efficace et exigeant l'élimination des mots fonctionnels (ang. *POS filters*), ensuite  $\chi^2$  (NKJP) et *log-likelihood* (Frantext) (Dryjańska 2020), les deux privilégiant l'extraction des collocations fréquentes, et puis, Information Mutuelle (Frantext) pour extraire des collocations moins fréquentes.

### **Sources potentielles d'erreurs dans l'analyse de corpus**

L'extraction de collocation fondée sur les outils automatiques a l'avantage d'analyser de grandes quantités de données, ce qui ne serait pas possible manuellement. Or, cette analyse présente des sources d'erreurs de types différents et dépendant d'un corpus.

Tout d'abord, la lemmatisation n'est pas toujours correcte, surtout en polonais, vu sa grande variété flexionnelle. Ensuite, l'outil de recherche de collocations dans NKJP et DLKZ ne prend pas en considération les signes de ponctuation et propose comme collocations des suites de mots séparés par la virgule, le point, le point d'interrogation ou le point d'exclamation. Puis, le classement des collocations est fondé sur le voisinage souvent trompeur, notamment en polonais dans lequel l'ordre des mots dans la phrase est flexible.

Exemple : *pragnący zabrać dziecko na święta muszą mieć zgodę sądu* 'ceux qui désirent emmener l'enfant à la fête doivent disposer d'une autorisation du tribunal'. Le verbe *musieć* 'devoir', se référant au sujet *pragnący*, n'est donc pas le collocatif de *święta*.

Les collocations formées par plus de deux mots ne sont pas affichées dans NKJP ; uniquement un élément d'une suite de plusieurs mots est proposé comme un collocatif. Frantext permet d'extraire des collocations qui contiennent trois éléments : nom – préposition – nom (NPN). Cependant, l'analyse automatique n'a pas permis d'extraire toute la suite de mots *fêtes de fin d'année*, mais il fallait ajuster « manuellement » le résultat de l'extraction automatique dans Frantext et Leipzig Corpora Collection French.

### **Étape préliminaire : analyse de *fête*, *święto* et *šventė* dans Leipzig Corpora Collection**

Leipzig Corpora Collection possédant une interface qui visualise les cooccurrences significatives est un outil facilement intégrable en didactique des langues. Les étudiants, presque sans connaissances préalables, peuvent voir et interpréter certaines cooccurrences fréquentes d'un lexème. Cela permettrait de commencer à développer une démarche *hands-on* en classe. Dans notre recherche, les graphes présentés par Leipzig Corpora Collection nous ont permis de balayer rapidement les

contextes importants des lexèmes *fête*, *święto* et *šventė* en cherchant des points de convergence et de divergence. Comme notre recherche consiste à construire l'image linguistique de *fête*, nous nous concentrons sur les collocatifs typants : (A)N(A) en trois langues, NN en polonais et lituanien, aussi bien que NdeN en français, qui figurent parmi les patrons énumérés par Hausmann (1989).

Les graphiques suivants représentent les cooccurrences significatives (les résultats présentés par les graphes sont des cooccurrences fréquentes délimitées par la phrase) de *fête* et *fêtes*, *święto* et *święta*, et *šventė*, *šventės*. Nous présentons aussi le nombre d'occurrences dans le corpus entier et la classe de fréquence (les lexèmes plus fréquents ont la classe de fréquence plus basse).

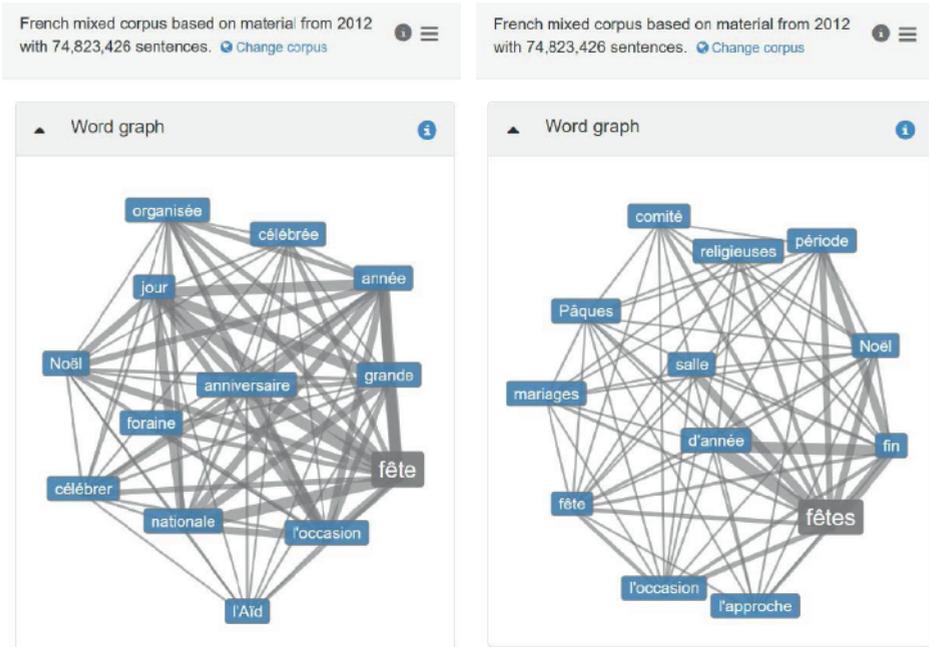


Figure 1. Analyse de *fête* dans Leipzig Corpora Collection

Mot : *fête* Nombre d'occurrences: 111 832 **Classe de fréquence: 9**

Mot : *fêtes* Nombre d'occurrences: 43 115 **Classe de fréquence: 11**

Tout d'abord, il est à noter que les nombres d'occurrences des deux formes ne sont pas proches. *Fête* est plus fréquente (111 832) que *fêtes* (43 115). Nous remarquons la divergence associative des formes – *fête* et *fêtes* ; il n'y a qu'une seule cooccurrence commune : *fête(s) de Noël*. Dans les autres cas, on distingue la différence des cooccurrences au singulier : *fête d'anniversaire*, *fête de l'Aid* (*fête musulmane*), *fête foraine*, *grande fête* et de celles au pluriel : *fêtes de fin d'année*, *fêtes religieuses*, *fêtes de Pâques*. La distinction entre les cooccurrences au singulier et au pluriel existe aussi en polonais : nous parlons de *Święta Bożego Narodzenia* 'fêtes de Noël' et *Święta Wielkanocne* 'fêtes pascales'.

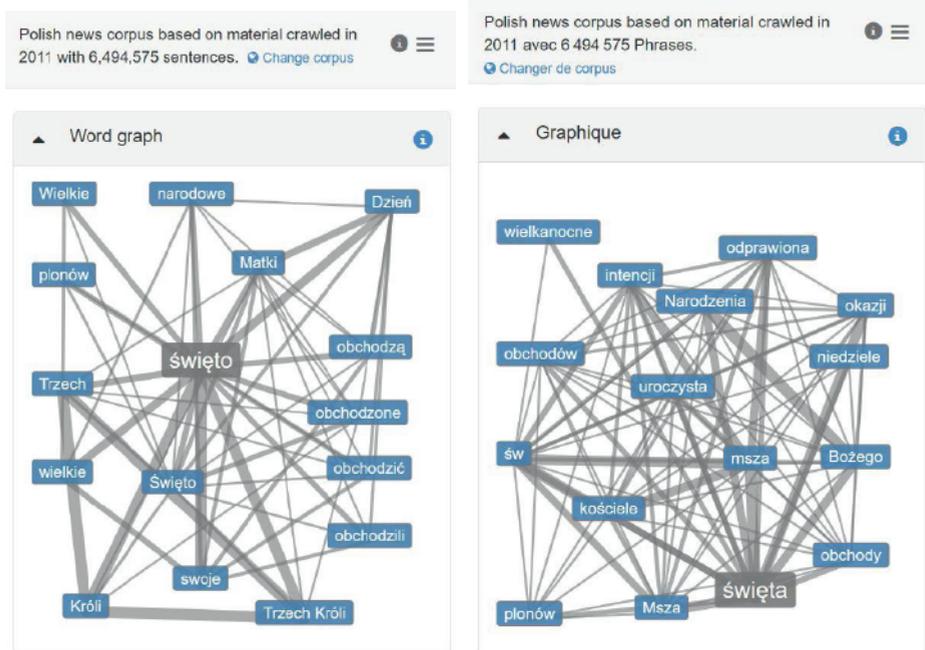


Figure 2. Analyse de *święto* dans Leipzig Corpora Collection

Mot : *święto* Nombre d'occurrences: 3,795 **Classe de fréquence : 10**

Mot : *święta* Nombre d'occurrences: 6,625 **Classe de fréquence : 9**

Parmi les cooccurrences significatives, il y a *święto narodowe* 'fête nationale', *święto plonów* 'fête de la moisson', *Święto Trzech Króli* 'fête des Rois Mages', *Święto Matki* 'fête de la Mère', *Święta Bożego Narodzenia* 'fêtes de Noël', *Święta Wielkanocne* 'fêtes pascales'.

L'analyse plus approfondie a révélé que *Święto Matki* constituait une partie d'une cooccurrence de trois mots – *Święto Matki Bożej* ou *Boskiej* 'fête de la Mère de Dieu'. Ce qui est pourtant intéressant c'est que, parmi les cooccurrences présentées dans la figure 2, il y a aussi une collocation *Dzień Matki* 'journée de la Mère'.

En ce qui concerne le sous-corpus lituanien, la situation est présentée dans la figure 3 :

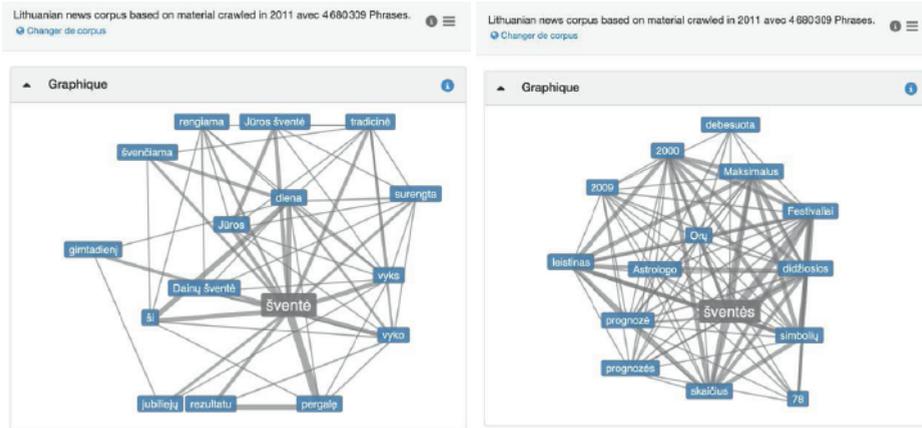


Figure 3. Analyse de *šventė* dans Leipzig Corpora Collection

Mot-clé: *šventė* (n. sing. féminin.). Nombre d'occurrences: 8 323. **Classe de fréquence: 8.**

Mot-clé: *šventės* (n. plur. féminin.). Nombre d'occurrences: 29 831. **Classe de fréquence: 6.**

Les collocations les plus fréquentes dans ce corpus sont : *Jūros šventė* 'fête de la Mer', *Dainų šventė* 'fête des chansons', *tradicinė šventė* 'fête traditionnelle', *miesto šventė* 'fêtes des villes', *žiemos šventė* 'fêtes d'hiver'. Tandis que la forme plurielle de *šventės* ne représente pas des collocations significatives, le seul cas à mentionner est la collocation *didžiosios šventės* 'les plus grandes fêtes' c.à.d. les plus importantes fêtes pour la nation lituanienne. La fête lituanienne est généralement déterminée par les coutumes, les traditions, d'autres normes de la société, les institutions internationales ou d'État.

## Étape 2 : analyse de *fête*, *święto* et *šventė* dans Frantext, NKJP et DLKT

A cette étape, nous avons relevé des cooccurrences en fonction de deux patrons :

- substantif – substantif (NN) en polonais et en lituanien et leur équivalent N(de) N en français ;
- substantif – adjectif (NA) et adjectif – substantif (AN) en trois langues.

Le lexème pol. *święto* avec toutes ses formes morphologiques, apparaît 26703 fois dans le corpus NKJP; cependant parmi ces formes, il y a celles qui représentent différentes catégories grammaticales. Une analyse correctrice a été effectuée, ce qui finalement a permis d'obtenir approximativement 20 000 occurrences du substantif *święto* étant l'ajustement de 27,84% par rapport à la valeur initiale. Dans NKJP, nous avons pu relever 628 collocations de type NN et 302 collocations de type (A)N(A).

Pour débiter l'analyse du corpus du lituanien contemporain *Dabartinės lietuvių kalbos tekstynas* (DLKT) une liste des mots du corpus a été constituée. Parmi toutes les sources proposées par celui-ci, le lexème *šventė* et ses formes fléchies *šventė* ~ *švente* ~ *Šventė* ~ *Švente* apparaissent 7251 fois. Après l'analyse quantitative effectuée manuellement, en rejetant les homofonies verbales (*šventė* 'fêtait' v. 3p.

sing.), les séquences non logiques, le corpus permet de relever le nombre de 2793 collocations.

Le lexème *fête* et toutes ses formes fléchies apparaissent dans Frantext 19243 fois, donc *fête* et *święto* ont la fréquence semblable. Tandis que le nombre d'occurrences du lexème *šventė* dans le corpus du lituanien contemporain est beaucoup plus restreint, en conséquence on a analysé parallèlement pour le lituanien les données de Sketch Engine avec son nombre de 10367 du lexème *šventė*, pour mettre en équilibre les données des corpus. Dans Frantext, le type de correction « manuelle » qui a été effectuée en polonais ou en lituanien n'était pas nécessaire, car le concordancier permet de préciser quelle catégorie grammaticale doit être prise en considération dans la recherche.

Frantext permet d'observer approximativement 8200 collocations de type N(de)N de *fête(s)* et 4750 collocations selon le type NA.

Les 10 collocations les plus répandues dans les corpus de trois langues révèlent deux types syntaxiques de collocations nominales qui prévalent, la construction de deux noms et un nom plus adjectif. Le premier tableau 1 présente la structure purement nominale. L'on peut y apercevoir quelques coïncidences de trois langues pour les collocations *fêtes de ville, de famille, de Pâques*, les tableaux 4–6 représentent pourtant une liste des collocations les plus répandues du type NA qui illustre plutôt une non-coïncidence entre les tableaux. Parmi les collocations fréquentes, l'on peut mentionner *fête religieuse, fête nationale et fête d'État*.

Les tableaux suivants présentent les 10 collocatifs les plus fréquents de chaque langue.

Tableau 1. Collocatifs nominaux N(de)N (type de fêtes) dans Frantext

Collocatifs nominaux de fête	Fréquence (Frantext)
(Fête)/de Dieu	195
famille	122
Noël	115
mère	113
Pâques	111
saint	110
village	99
mort	49
père	45
printemps	37

Tableau 2. Collocatifs nominaux de *święto* dans les collocations de type NN dans NKJP

Collocatifs nominaux de <i>święto</i>		Fréquence (NKJP)
indépendance	<i>niepodległość</i>	518
ville	<i>miasto</i>	288
moisson	<i>plon</i>	233

mort	<i>zmarły</i>	224
mère de Dieu	<i>Matka Boża</i>	141
police	<i>policja</i>	139
travail	<i>praca</i>	129
famille	<i>rodzina</i>	114
armée	<i>wojsko</i>	107
Pâques	<i>Wielkanoc</i>	104

Tableau 3. Collocatifs nominaux de šventė dans les collocations de type NN dans DLKT

Collocation NN (LT)		Fréquence (DLKT)
Fête de la mer	<i>Jūros šventė</i>	CLEI 4668 DLKT 46 SE 240
Fête des chansons	<i>Dainų šventė</i>	CLEI 1576 DLKT 111 SE 1149
Fête de sport	<i>Sporto šventė</i>	CLEI 390 SE 123 DLKT 75
Fête de ville	<i>Miesto šventė</i>	CLEI 307 SE 208 DLKT 64
Fête de la Saint-Jean	<i>Joninių šventė</i>	CLEI 328 SE 49 DLKT 9
Fête de rosée	<i>Rasos šventė</i>	SE 145 DLKT 18 CLEI 242
Fête de Lituanie	<i>Lietuvos šventė</i>	CLEI 183 SE 268 DLKT 5
Fête de famille	<i>Šeimoms šventė</i>	CLEI 268 SE 552 DLKT 7
Fête d'hiver	<i>Žiemos šventė(s)</i>	CLEI 285 SE 107 LKT 11
Fête de la moisson	<i>Derliaus šventė</i>	CLEI 291 SE 84 DLKT 0

Tableau 4. Collocatifs adjectivaux de fête dans les collocations de type NA dans Frantext

Collocatifs adjectivaux de fête	Fréquence (Frantext)
forain	221
national	188

public	126
religieux	119
patronal	47
populaire	46
champêtre	37
local	24
juif	22
familial	19

Tableau 5. Collocatifs adjectivaux de *święto* dans les collocations de type (A)N(A) dans NKJP

Collocatifs adjectivaux de <i>święto</i>		Fréquence NKJP
de Noël	<i>boży (Boże Narodzenie)</i>	1775
Pascal	<i>wielkanocny</i>	639
national	<i>narodowy</i>	470
d'Etat	<i>państwowy</i>	234
de l'église	<i>kościelny</i>	137
religieux	<i>religijny</i>	121
familial	<i>rodzinny</i>	120
polonais	<i>polski</i>	117
de mois de mai	<i>majowy</i>	83
du peuple	<i>ludowy</i>	74

Tableau 6. Collocatifs adjectivaux de *šventė* dans les collocations de type (A)N(A) dans DLKT

Collocatifs adjectivaux de <i>šventė</i>		Fréquence DLKT
traditionnelle	<i>tradicinė</i>	CLEI 252 SE 49 DLKT 15
d'anniversaire, commémorative	<i>jubiliejinė</i>	CLEI 216 DLKT 2
républicaine	<i>respublikinė</i>	CLEI 211 DLKT 2
professionnelle	<i>profesinė</i>	CLEI 179 DLKT10
d'État	<i>valstybinė</i>	CLEI 177 SE 113 DLKT 27
lituanienne	<i>lietuvių</i>	SE 248 DLKT 0
liturgique	<i>liturginė</i>	CLEI 152 DLKT 6

Nationale	<i>Nacionalinė</i>	CELI 83 SE 64 DLKT 15
religieuse	<i>religinė</i>	CLEI 70 SE 34 DLKT 22
païenne	<i>pagoniška</i>	CLEI 10 DLKT 6

En analysant les collocations N(de)N et (A)N(A) les plus fréquentes en français, en polonais et en lituanien, nous avons constaté la présence de deux collocations communes *fête religieuse* et *fête nationale*. De plus, on observe plus de similitudes entre deux langues. Pour le français et le polonais, il faut évoquer les collocations similaires – *fête familiale*, *fête de famille*, *fête pascalle*, *fête de Pâques* et *fête de Noël*. Quant au polonais et lituanien, on observe les collocations semblables suivantes : *fête d'État*, *fête lituanienne* ou *polonaise*, *fête de la ville*, *fête de la moisson*.

Dans les langues, la même présentation sémantique de la *fête* peut se réaliser non seulement par différentes structures syntaxiques, mais aussi par d'autres mots : *journée*, *festival* ou parfois *mois*, *semaine*, etc., ce qui est pourtant l'objet d'une autre recherche.

Cette analyse, permettant d'observer la fréquence de différentes collocations, constitue un enrichissement par rapport aux informations lexicographiques et par conséquent elle pourrait être intéressante pour les apprenants de français, surtout dans le cadre philologique. Il est à noter que les apprenants d'aujourd'hui en tant qu'utilisateurs d'Internet, habitués au critère de fréquence accompagnant leurs requêtes dans des moteurs de recherche, posent souvent eux-mêmes des questions concernant la fréquence d'emploi de mots ou d'expressions. Jusqu'à l'avènement des grands corpus de texte, les réponses à ces questions étaient basées sur l'intuition de l'enseignant. Quelque profonde qu'elle soit, elle est rarement fondée sur une aussi grande diversité de contextes offerte par les grands corpus comme Frantext, NKJP ou DLKT.

### Catégorisation collocationnelle

Afin d'atteindre l'objectif interculturel de notre recherche, nous procédons à une catégorisation de toutes collocations relevées, à savoir N(de)N et (A)N(A). Initialement, nous avons voulu distinguer les mêmes catégories pour les trois langues pour une analyse quantitative, ce qui ne s'est avéré que partiellement possible. Les graphiques suivants présentent le résultat de la catégorisation en français, en polonais et en lituanien ; les trois premiers pour chaque langue séparément et le quatrième est un graphique contrastif représentant toutes les catégories dans les trois langues ensemble. Toutes ces catégories peuvent être regroupées dans deux *supracatégories* : fêtes religieuses et fêtes laïques.

### Types de fêtes – analyse des collocations sur Frantext

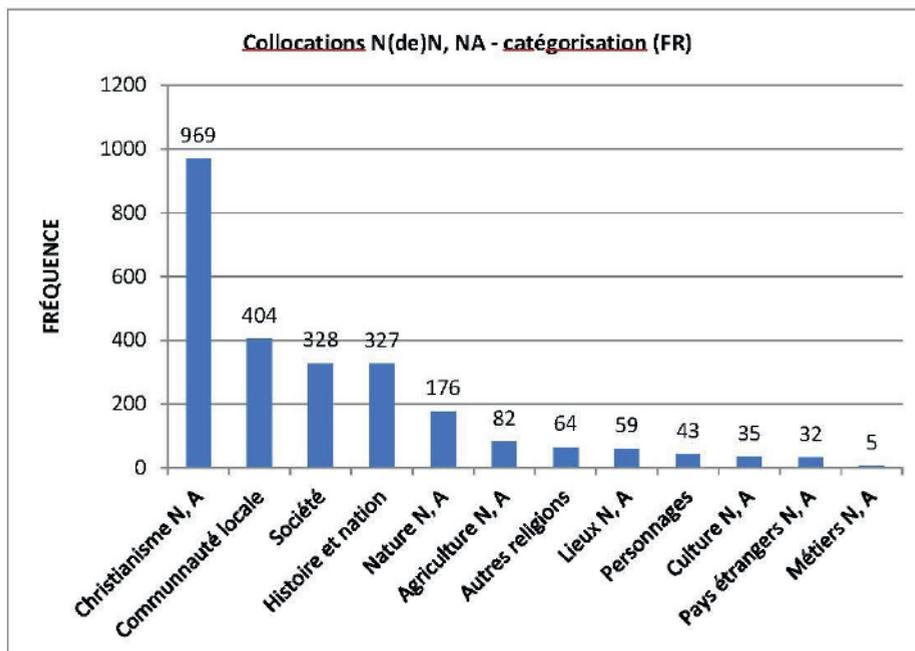


Figure 3. Les collocatifs nominaux et adjectivaux de *fête* à caractère typant catégorisés en 12 catégories (Frantext)

Pour le lexème *fête*, nous avons distingué douze catégories qui contenaient les collocations comme suit :

Les fêtes liées à différentes traditions religieuses ou cultes:

- Christianisme – fête-/de Dieu, fête de Noël, ...
- Autres religions – fête juive, fête païenne, ...

Les fêtes laïques :

- Histoire et nation – fête nationale, fête du 14 juillet, ...
- Communauté locale – fête foraine, fête du village, ...
- Lieux – fête de Paris, fête de Versailles, fête de Neuilly, ...
- Société – fête des Mères, des Pères, ...
- Nature – fête du printemps, fête du soleil, ...
- Agriculture – fête champêtre, fête de la moisson, ...
- Personnages – fête de Bacchus, fête d’Adonis, fête de Vénus, ...
- Culture – fête de l’école, fête du carnaval, ...

L’apprenant, tout conscient qu’il soit de l’importance de la laïcité en France, peut être étonné par la dominance de la catégorie Christianisme. Il s’avère que la tradition chrétienne est très ancrée dans la langue. Ensuite, nous observons un équilibre approximatif des trois catégories : Communauté locale, Société et Histoire et nation.

## Types de fêtes – analyse des collocations dans NKJP

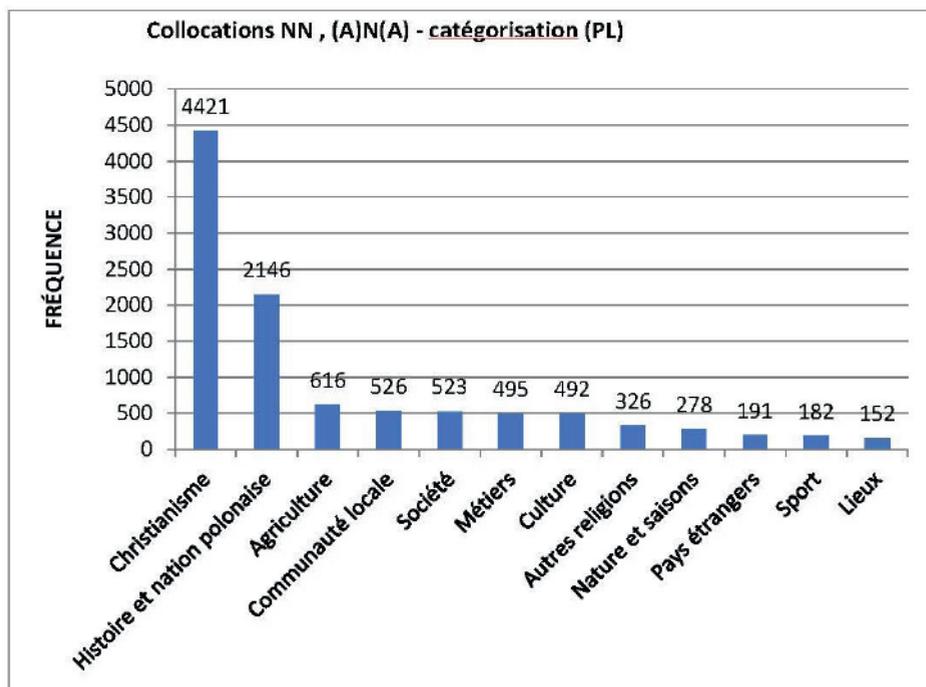


Figure 4. Les collocatifs nominaux et adjectivaux de *święto* à caractère typant catégorisés en 12 catégories (NKJP)

En polonais, nous avons aussi distingué douze catégories, dont la plupart sont celles vues pour le français. On note pourtant deux différences : une nouvelle catégorie Sport, l'absence de la catégorie Personnages.

Tout d'abord, les catégories les plus fréquentes en polonais sont Christianisme et Histoire et nation, relation explicable par une histoire polonaise très difficile. C'est l'histoire d'une lutte acharnée, soutenue par l'Église catholique, de conserver la *polonité*. Souvent les fêtes religieuses étaient l'occasion de manifester librement l'attachement à la culture, la tradition et la langue polonaises. La mémoire collective des événements historiques polonais, glorieux ou tragiques, est toujours vivante d'où un grand besoin de les commémorer, ce qui explique la force de la catégorie Histoire et nation et la présence des collocations comme *fête nationale*, mais également *fête de l'armée polonaise*, fête de la Constitution, fête du 3 mai. Notons aussi que la dominance de ces deux catégories correspond aux résultats d'une recherche semblable sur le lexème *dziedzictwo* (fr. *patrimoine*) (Dryjańska 2020) qui a démontré que ses collocations les plus fortes étaient *patrimoine national* et *patrimoine chrétien*.

Le graphique ci-dessus permet ensuite d'observer un quasi-équilibre entre les cinq catégories : Agriculture, Communauté locale, Société, Métiers et Culture, et un autre groupe de catégories équilibrées : Autres religions, Nature, Pays étrangers, Sport et Lieux.

Si l'on compare la distribution de ces catégories à celles en français, on est frappé par une observation : la force des catégories Communauté locale et Société est comparable à celle de la catégorie Histoire et nation en français, ce qui n'est pas le cas en polonais. Un équilibre visible entre des festivités françaises liées à la cause nationale et à la vie d'une communauté locale versus un fort accent sur les aspects nationaux au détriment des commémorations locales polonaises constitue une caractéristique différenciant les deux pays.

### Analyse de *šventė*

L'éventail des collocations lituaniennes à la base de *šventė* dans le corpus de Sketch Engine diffère par sa plus grande attention à la nation, l'on a donc proposé même deux sous-catégories de l'histoire et de la nation lituanienne: celle qui est liée à l'État, au gouvernement (*valstybės* 'd'Etat', *nepriklausomybės* 'de l'Indépendance') et autre qui est liée plutôt aux traditions du peuple comme *dainų šventė* 'fête des chansons folkloriques', *Rasos šventė* 'fête de rosée' etc. Les mêmes catégories qui ont les fréquences comparables sont plus ou moins fortement fréquentes dans les autres corpus lituaniens analysés. Le graphique ci-dessous illustre les données de trois corpus lituaniens :

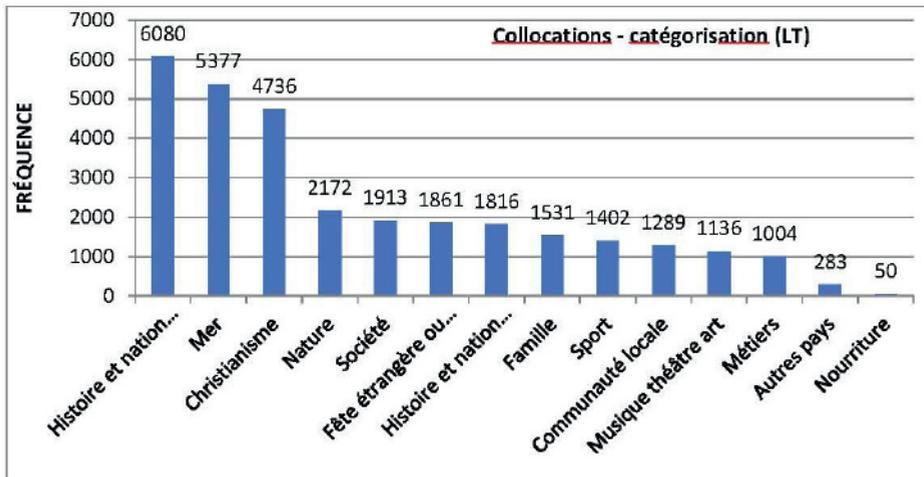


Figure 5. Les collocatifs nominaux et adjectivaux de *šventė* à caractère typant catégorisés en 14 catégories

Dans le graphique 5, entre autres, la catégorie « Mer » (lituanien *Jūra*) se distingue nettement. Un taux élevé de cette catégorie, même si elle n'est pas très variée par son inventaire, s'explique plutôt par la fréquence d'une collocation *Jūros šventė* 'la fête de la mer' qui est très forte à son tour. Une autre catégorie historiquement importante est liée au christianisme. Ces données s'expliquent par le fait que les jours fériés lituaniens ont été influencés par une vision du monde païenne, puis chrétienne, dont certains sont devenus partie intégrante d'un culte religieux.

Après le baptême de la Lituanie (1387), les jours fériés étaient règlementés par l'Église catholique. Cependant, l'histoire lituanienne est aussi marquée par des

périodes où les fêtes religieuses étaient interdites sur le territoire tandis que les festivals de chants traditionnels et les festivals de récolte ont survécu. Ces moments historiques pourraient être pris en considération en commentant les chiffres de la figure 5 afin d'expliquer la fréquence de certaines catégories et ses différences avec les données polonaises. Toutefois, non seulement les faits historiques et l'aspect interculturel pourraient être la raison d'une certaine image statistique, mais aussi des particularités linguistiques d'une langue. Ainsi, en polonais et en français, il s'agit de collocations telles que *la fête de Noël* et *la fête de Pâques*, tandis qu'en lituanien ce n'est qu'un seul mot qui nomme la fête : *Kalėdos* 'Noël' et *Vėlykos* 'Pâques'. Donc cette différence syntaxique influence à son tour, elle aussi, cette non-coïncidence collocationnelle entre les langues et, par conséquent, elle devrait être éclaircie en classe de langue.

Les liens entre la communauté rurale traditionnelle et l'importance de la culture rurale se manifestent dans la catégorie de la nature aussi bien que de la nourriture, il s'agit des collocations telles que *derliaus šventė* 'fête de la bonne récolte', *grybų šventė* 'fête des champignons', *alaus šventė* 'fête de la bière', etc.

Les résultats des graphiques 3 – 5 représentent une preuve indéniable de l'importance de la compétence interculturelle en classe, grâce auxquels les apprenants découvrent l'histoire, la culture, la tradition d'un pays en apprenant le champ lexical, la syntaxe et la morphologie. Cette complexité des collocations révèle la nature d'une langue qui est le miroir d'une société. D'après Binon et Verlinde (2003 : 31) la maîtrise des collocations constitue la clef de voûte de l'enseignement et de l'apprentissage d'une langue seconde, surtout au niveau intermédiaire. Une analyse de la fréquence des collocations démontre d'ailleurs que leur fréquence dépasse souvent celle de mots tout à fait courants.

Ensuite, nous avons voulu comparer la « force » de chaque catégorie dans toutes les langues, ce qui a été montré dans le graphique ci-dessous.

La figure 6 illustre que la catégorie dominante en français et en polonais Christianisme est plus forte pour *święto*. Ensuite, nous observons la dominance de la catégorie Histoire et nation en lituanien et en polonais, à l'avantage du lituanien. Puis, l'on remarque l'importance de la catégorie Société, qui est surtout visible pour le français, et aussi pour le lituanien. On distingue aussi une grande dominance de la catégorie Communauté locale uniquement en français.

Pour comparer quantitativement toutes ces catégories collocationnelles, nous avons calculé le coefficient de corrélation linéaire pour toutes les paires des langues analysées, ce qui a conduit aux résultats suivants :

- coefficient de corrélation linéaire fr vs pl – 0,9 (très haut) ;
- coefficient de corrélation linéaire lt vs pl – 0,54 (les résultats des 3 corpus lituaniens) ;
- coefficient de corrélation linéaire fr vs lt – 0,45 (les résultats des 3 corpus lituaniens).

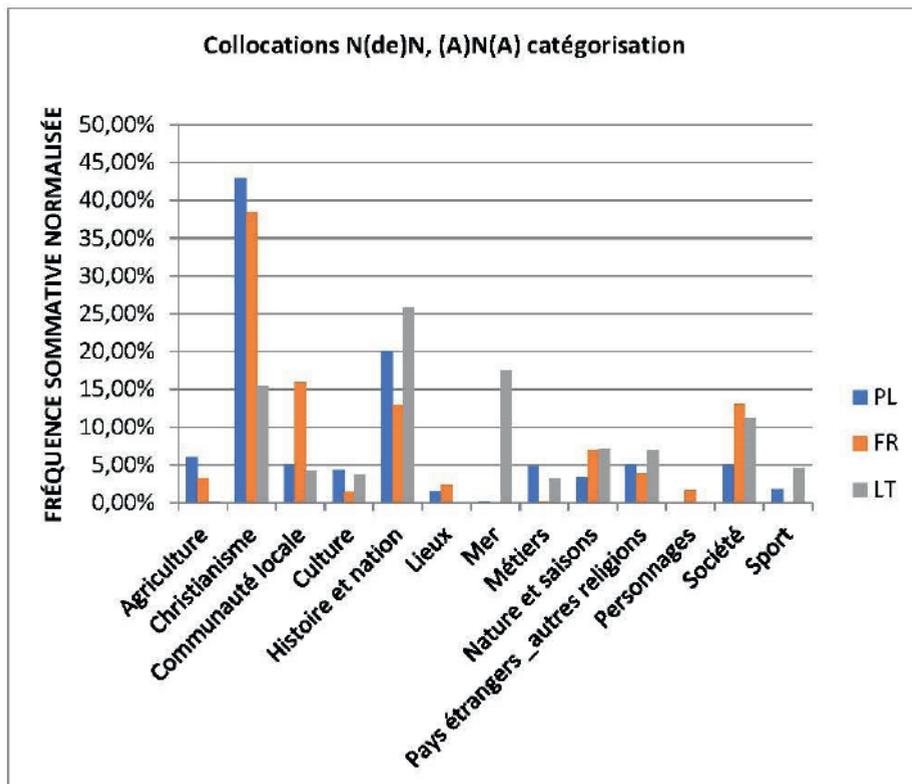


Figure 6. Analyse contrastive des catégories collocationnelles normalisées en fr, pl et lt

Ce qui est étonnant, c’est une basse corrélation de la distribution des catégories polonaises et des catégories lituanienne. Rappelons que l’une des hypothèses de cette recherche était une coïncidence collocationnelle des mots *święto* et *šventė* fondée sur la proximité des cultures polonaise et lituanienne. Or, le calcul des coefficients de corrélation a démontré une relation opposée : la corrélation est plus forte entre la distribution des catégories collocationnelle de *fête* et de *święto*, ce qui peut résulter d’une ressemblance des phénomènes festifs en France et en Pologne. Il est pourtant à noter que le résultat concernant les catégories lituanienne peut être biaisé par l’incompatibilité partielle de corpus lituanien par rapport aux corpus français et polonais.

### Étape trois : analyse complémentaire fondée sur Sketch Engine

Tableau 7. Les collocations les plus fréquentes de *fête* relevées de Sketch Engine

Collocations de <i>fête</i>	fréquence
Fête nationale	10275
Fête de la musique	8517
Fête foraine	5297
Fête de Noël	3612

Joyeuse fête	2828
Fête patronale	2088
Fête de (différents) saints	2072
Fête populaire	1940
Fête religieuse	1800
Fête de Pâques	1617
Fête de village, du village	1387
Fête des voisins	1379
Fête-/Dieu	1378
Fête de quartier	833
Fête communale	444
Fête publique	165

Les données du tableau 7 confirment en grande partie les résultats de la recherche dans Frantext : des collocations importantes sont les suivantes : *fête nationale, foraine, de Noël, de Pâques, du village, de quartier, communale, populaire*. Or, elle révèle les collocations comme *fête de la musique* et *fête des voisins* – les fêtes plus récentes donc observables en tant que collocations dans le corpus fondé sur les ressources Internet. Notons également la fréquence plus basse de *fête foraine* que celle de *fête nationale* et la fréquence plus basse de *Fête-/de Dieu* que celle de *fête de Noël* ou de *Pâques*. Ces observations montrent l'importance du choix de corpus que l'on intègre à la didactique du FLE. Le corpus comprenant des textes plus récents permet de dresser l'image linguistique d'un mot plus complète. Cela permet aussi d'observer une évolution dans le temps de la fréquence de collocations. L'intérêt pour certaines célébrations change, ce qui implique le nombre de cooccurrences dans des corpus.

## Conclusions

Dans cette étude, nous n'avons pas prétendu que toutes les expressions lexicalisées et compositionnelles pouvaient être analysées comme des collocations. Nous avons souhaité montrer que, parmi différents types d'expressions lexicalisées, il existait un schéma sémantique et syntaxique particulièrement propice à la lexicalisation et à la spécialisation du sens de chaque langue. Et c'est grâce aux corpus à notre disposition<sup>5</sup> que nous avons extrait certaines caractéristiques des collocations lexicales propres à trois langues. Les conclusions présentent une synthèse des principaux résultats.

1. Une grande dominance de la catégorie Christianisme (surtout en fr et pl), due aux racines chrétiennes communes, les caractéristiques des autres catégories distinguées étant plus hétérogènes.

<sup>5</sup> Ajoutons pourtant qu'il serait intéressant de continuer cette analyse si l'on disposait de corpus plus compatibles dans les trois langues dans le futur en vue d'une vérification de nos résultats actuels.

2. La présence des catégories propres à une ou l'autre langue, par ex : une très forte catégorie Histoire et nation en lituanien comprenant deux sous-catégories et une forte catégorie Mer.
3. Contrairement aux attentes, il paraît que les images linguistiques de la fête en français et en polonais sont plus proches (corrélation haute, homogénéité des catégories) qu'entre le lituanien et le polonais.
4. Parmi les collocations relevées, la structure N(de)N prévaut dans les trois langues.
5. L'analyse, certainement non exhaustive, indique la grande variété de la non-coïncidence collocationnelle du mot *fête* (fr, lt, pl) qui doit être prise en compte dans l'élaboration des méthodes et des démarches en classe afin de les adapter aux particularités linguistiques et culturelles des apprenants.
6. Il y a une possibilité de continuer et de compléter notre étude pour découvrir le sens de la fête par l'analyse collocationnelle des synonymes du mot *fête* en français, en lituanien et en polonais.

Il serait nécessaire de systématiser l'enseignement/apprentissage des collocations auprès des apprenants en FLE. Il s'agirait d'aider les élèves à retenir le sens et l'association lexicale simultanément en FLE, ainsi que de fixer les structures qu'ils connaissent déjà partiellement et de découvrir l'aspect culturel. Ce type de données pourrait en outre constituer une base pour diverses démarches scientifiques relatives à l'apprentissage des langues étrangères.

## Bibliographie

- Anctil D., Tremblay O., 2016, *Les collocations : des combinaisons de mots privilégiées, Dictionnaire Lexique*, „Logiciels d'aide à la rédaction” XXI (3).
- Benigno V., Grossmann F., Kraif O., 2015, *Les collocations fondamentales : une piste pour l'apprentissage lexical*, „Revue française de linguistique appliquée 2015” XX (1), pp. 81–96.
- Binon J., Verlinde S., 2003, *Les collocations : clef de voûte de l'enseignement et de l'apprentissage du vocabulaire d'une langue étrangère ou seconde*, „La Lettre de l'AIRDF”, XXXIII (2), pp. 31–36.
- Bouma G., 2009, *Normalized (Pointwise) Mutual Information in Collocation Extraction*. „Proceedings of GSCL”, Department Linguistik, Universität Potsdam.
- Cholewa J., 2008, *Image encyclopédique et linguistique du chat et du chien en français et en polonais contemporains*, Wydawnictwo Uniwersytetu w Białymstoku, Białystok.
- Długosz-Kurczabowa K., 2004, *Jaka jest genealogia słów święto, świętować?*, PWN, <https://sjp.pwn.pl/poradnia/haslo/genealogia-swiat;5824.html> (consulté le 25 février 2022).
- Dryjańska A., 2020, *Patrimoine vs. dziedzictwo – interculturality in French language teaching*, „Text und diskurs” XIII, pp. 175–193.
- Evert S., 2007, *Corpora and collocations*. Osnabrück, Institute of Cognitive Science, University of Osnabrück.
- Fabre C., 2015, *Sémantique distributionnelle automatique : la proximité distributionnelle comme mode d'accès au sens*, „Éla. Études de linguistique appliquée” 2015/4 (N° 180), pp. 395–405.

- Firth J.R., Haas W., Halliday M.A.K., Allen W.S., Robins R.H., 1962 (1957), *Studies of linguistic analysis*, Basil Blackwell, Oxford.
- Halliday M.A.K., 1985, *An introduction to functional grammar*, London, Arnold.
- Hausmann F.J., 1989, *Le dictionnaire de collocations*, [dans:] *Dictionnaires. Encyclopédie internationale de lexicographie*, (Manuel de linguistique et des sciences de communication, 5/1), éd. F.J. Hausmann, O. Reichmann, H.E. Wiegand, L. Zgusta, Berlin/New York : De Gruyter, pp. 1010–1019.
- Hausmann F.J., 2007, *Lexicographie française et phraséologie*, [dans:] *Collocations, phraséologie, lexicographie. Etudes 1977–2007*, éd. E. Haag, Aachen : Shaker Verlag, pp. 121–153.
- Kazlauskienė V., 2022, *Collocations lexicales et leur acquisition en FLE*, „Sustainable Multilingualism Biannual Scientific Journal”, Vytautas Magnus University.
- Kilgarrriff A., 2001, *Comparing corpora*, „International Journal of Corpus Linguistics” XI (1), pp. 1–37.
- Manning C., Schütze H., 1999, *Foundations of Statistical Natural Language Processing*, Cambridge, MA : MIT Press.
- McEnery T., Wilson A., 2001 (1996), *Corpus Linguistics*, Edinburgh University Press, Edinburgh.
- Mel’čuk I., 1998, *Collocations and Lexical Functions*, [dans:] *Phraseology. Theory, Analyses, and Applications*, ed. A. Cowie Oxford: Oxford University Press, pp. 23–53.
- Mel’čuk I., Polguère A., 2007, *Lexique actif du français : l’apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*, Bruxelles : De Boeck.
- Mikolov T., Yih W., Zweig G., 2013, *Linguistic Regularities in Continuous Space Word Representations*, „HLT-NAACL”, pp. 746–751.
- Razauskas D., 2016, *Lietuviškosios šventumo sąvokos kilmė*, „Mokslo darbai. prigimtinės kultūros seminaras”, pp. 13–18.
- Rodríguez-Fernández S., Espinosa-Anke L., Carlini R., Wanner L., 2016, *Semantics-Driven Recognition of Collocations Using Word Embeddings. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, „Association for Computational Linguistics”, pp. 499–505.
- Royle P., Richardson I., Boisvert S., Bourguignon N., 2009, *Is Bigger Better? Corpus and Dictionary Use in the Search for Compounds, Collocations, Derived Forms and Fixed Expressions*, „Journal des traducteurs Translators” 54(3), pp. 520–532.
- Siepmann D., 2006, *Collocations et dictionnaires d’apprentissage onomasiologiques bilingues : questions aux théoriciens et pistes pour l’avenir*, „Langue française 107”, pp. 99–118.
- Silberztein M., 2015, *Formalisations des langues*, L’approche de NooJ. ISTE Éditions, London.
- Sinclair J., 1991, *Corpus, Concordance, Collocation*, Oxford University Press, Oxford.
- Śliwa D., 2011, *Coeur et sentiments: une étude cognitive des collocations françaises et polonaises*, [dans:] *Parler des émotions : entre langue et littérature*, dir. A. Krzyżanowska, R. Jakubczuk, Lublin, Wyd. UMCS, pp. 191–199.
- Stubbs M., 1996, *Text and Corpus Analysis: Computer Assisted Studies of Language and Culture*, Blackwell, Oxford.

Le sens de *fête* en polonais, en lituanien, en français et sa (non)coïncidence collocationnelle [41]

Teubert W., 2009, *La linguistique de corpus : une alternative*, „Critical Discourse Analysis I. Les notions de contexte et d'acteurs sociaux”, SEMEN.

Thomas J., 2015, *Discover English with Sketch Engine*, „A corpus-based approach”, Versatile.

Tognini-Bonelli E., 2001, *Corpus Linguistics at Work*, Amsterdam : Benjamin.

Tremblay O., 2014, *Les collocations : des mots qui font la paire*, „Québec Linguistique Appliquée” VII(1), pp. 7-26.

Tutin A., 2013, *Les collocations lexicales : une relation essentiellement binaire définie par la relation prédicat-argument*, „Langages” I(189), pp. 47-63.

Tutin A., Grossmann F., 2002, *Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif*, „Revue française de linguistique appliquée” 2002/I(7), pp. 7-25.

Wang H., Zou Y., 2019, *A Corpus-Based Study of Semantic Collocations of the Verb "Feel" in English Public Speaking Setting: Chinese EFL V.S Native Speakers*, „International Journal of English Linguistics” IX(1), pp. 251-260.

Xiao R., McEnery T., 2006, *Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective*, „Applied Linguistics” 27(1), pp. 103-129.

## Méthodes

Cathérine H., V. M. Kizirian, M. Waendendries, A. Berthet, E. Daill, 2012, *Alter ego + (1-5)*, Hachette FLE.

Alcaraz M., C. Braud, A. Calvez, G. Cornuau, A. Jacob, C. Pinson, S. Vidal, 2016, *Edito (A1-B2)*, Didier.

Mérieux R., Y. Loiseau, 2008, *Latitudes (1-3)*, Didier.

Deniset H., M. Capouet, 2021, *Sésame (1-2)*, Hachette FLE.

Girardet J., J. Pécheur, C. Gibbe, M.-L. Parizet, 2016, *Tendances (A1-B1)*, CLE international.

Cocton M.-N., D. Dupleix, E. Heu, E. Kasazian, D. Ripaud, 2016, *Saison (1-3)*, Didier.

## Dictionnaires

Antidote 8, 2018 *Druide* informatique inc.

TLFi – Trésor de la langue française informatisé, republié sur le site Centre National de Ressources Textuelles et Lexicales (CNRTL), <https://www.cnrtl.fr/definition/> (consulté le 13 mars 2022).

Doroszewski W. (red.), *Słownik języka polskiego*, PWN, <https://sjp.pwn.pl/doroszewski/lista> (consulté le 13 février 2022).

## Corpus

Dabartinės lietuvių kalbos tekstyną (DLKT), <http://tekstynas.vdu.lt/tekstynas/> (consulté le 13 mars 2022) (consulté le 13 mars 2022).

Frantext, Analyse et Traitement Informatique de la Langue Française (ATILF), <https://www.frantext.fr/> (consulté le 13 mars 2022).

Leipzig Corpora Collection, [https://corpora.uni-leipzig.de/en?corpusId=ind\\_mixed\\_2013](https://corpora.uni-leipzig.de/en?corpusId=ind_mixed_2013) (consulté le 3 avril 2022).

NKJP – Narodowy Korpus Języka Polskiego, <http://nkjp.pl/> (consulté le 13 mars 2022).

Sketch Engine <https://www.sketchengine.eu> (consulté le 6 avril 2022).

### **The meaning of *holiday* in Polish, Lithuanian, French and its collocational (non)coincidence**

The linguistic overview of the word *holiday* in the three languages (French, Lithuanian and Polish) is promising for the intercultural approach to teaching French as a foreign language with a view to go beyond the roughly monocultural contexts in Poland and Lithuania. The research is based on text corpora in these three languages. Its objective is to analyse the linguistic images of the word *holiday* and its Lithuanian and Polish equivalents and to examine their collocational (non)coincidence in order to systematize the teaching/learning of collocations to French learners. The aim would be to help students retain meaning and lexical association simultaneously, as well as to fix the structures they already partially know and to discover (inter)cultural aspects.

### **Znaczenie wyrazu *święto* w języku polskim, litewskim i francuskim oraz jego kolokacyjna (nie)ekwiwalentność**

Przedmiotem badania jest analiza kontrastywna wyrazu *święto* w trzech językach (francuskim, litewskim i polskim), której celem jest rozwijanie podejścia interkulturowego w nauczaniu/uczeniu się języka francuskiego w *grosso modo* monokulturowym kontekście edukacyjnym w Polsce i na Litwie. W analizie wykorzystana jest metodologia badań korpusowych, której kluczowym elementem jest kategoryzacja uzyskanych kolokacji.

Wykazała ona wiele różnic semantycznych, składniowych i frekwencyjnych dotyczących badanych leksemów, których odkrywanie powinno pomóc studentom w rozwijaniu kompetencji leksykalnej i interkulturowej.