



HAL
open science

Tombé en amour avec les corpus ?

Inka Wissner

► **To cite this version:**

Inka Wissner. Tombé en amour avec les corpus ?. Henry Tyne; Virginie André; Christophe Benzitoun; Alex Boulton; Yan Greub. French through Corpora: Ecological and Data-driven Perspectives in French Language Studies, Cambridge Scholars Publishing, pp.46-66, 2014, 978-1-4438-5831-1. hal-03607320

HAL Id: hal-03607320

<https://hal.science/hal-03607320v1>

Submitted on 16 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

CHAPTER XX

TOMBÉ EN AMOUR AVEC LES CORPUS? WHAT TEXT CORPORA CAN ADD TO THE STUDY OF DIATOPICALLY MARKED EXPRESSIONS OF MODERN FRENCH

Inka Wissner¹
ATILF research unit, France

1. Introduction

In modern French dialectology, the appropriate description of diatopic (i.e. regional) elements of the lexis of a national or regional variety (or of several dialects) includes identifying their distribution in time and space (cf. Rézeau 2007). Traditionally, researchers have mainly used lexicographic resources. However, these are based upon heterogeneous discourse genres. In addition, they do not describe all parts of the lexis systematically. Moreover, dictionaries specialising in phraseology hardly take into account the diatopic dimension of the language they describe (or do not even do so at all).

For the study of diatopically marked expressions it is, therefore, essential to search ecological data in text collections. The latter are seen as standard tools in modern French dialectology (e.g. Vézina 1998: 228).

¹ I wish to thank Myriam Bergeron-Maguire for comments on an early version of this text. My work has also benefited from exchanges with Jean-Paul Chauveau and André Thibault. I am grateful to the latter for giving me access to his corpus (ZOBEL), to Andres Kristol and Anne-Catherine Simon for having consulted SUISTEXT and VALIBEL respectively on my behalf, as well as to Alexandre Grandmaître from the Canadian Parliament's publication services, for having offered access to Hansard's non-public integral corpus.

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

However, this does not mean that their use is systematic or widespread. In fact, linguists encounter various problems, from practical access to resources to limited possibilities for critical philological text analysis (cf. Cappeau & Gadet 2007). In addition, many corpora (untagged) do not permit the retrieval of lemmatised sequences; also, they do not generally offer dialectal annotation, and are mostly not conceived for identifying the diatopic dimension of French (Wissner 2012a).

Bearing this situation in mind, I will consider how corpora help linguists to study phraseological diatopicisms in modern French. I focus on the expression *tomber en amour (avec quelqu'un)*, which is a near equivalent to Modern French *tomber amoureux (de quelqu'un)*. It is largely conceived as an emblematic regionalism of Quebec French, and as a calque of the English counterpart *to fall in love* (e.g. Parmentier 2006: 11). However, can we be sure that this French phrase is only used in varieties where its usage can be explained by English influence? If so, what do the available data allow us to interpret in terms of its usage and its history?

This chapter first looks at what the main French dictionaries tell us about the expression, starting with the *Trésor de la langue française* (TLF) and the *Französisches Etymologisches Wörterbuch* (FEW). It also explores other general French dictionaries from Estienne's *Dictionnaire Françoislatin* to the *Petit Robert* (PR), as well as complementary dictionaries on diatopicisms, up to the recent Belgian and Louisiana French dictionaries, *Dictionnaire des belgicisms* (DictBelg) and *Dictionary of Louisiana French* (DLF).

The chapter then examines occurrences of this expression in large text corpora, concentrating on the main collections of Modern French in the northern part of the Francophone world. The term *corpus* is here used in a large sense, referring to any collection of speech productions, consisting of integral texts or of text extracts, and conceived for scientific analysis (according to specific selective criteria), or for other purposes. The chapter goes on to explore traditional collections, mainly constituted of literary data (such as FRANTEXT), journalistic corpora (such as EUREKA), transcriptions of spoken data (such as in the PFC corpus), as well as text collections from the Internet. Amongst the latter, web-based FRWAC (cf. Baroni *et al.* 2009) is currently the largest monolingual French corpus freely available that has been constituted for research purposes (see

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

Wissner 2012a for an overview on existing corpora for studying French diatopicisms).

In order to analyse the phraseological diatopicism, I will consider the pertinent aspects of its occurrences in every checked resource. This includes its association with English or with Quebec French; its semantic, formal and variational characteristics, mainly diachronic and diatopic; its degree of lexicalisation; as well as its syntagmatic and paradigmatic behaviour. In the first instance I use the term *expression* (sometimes *phrase*) (see following section) to refer to *tomber en amour (avec quelqu'un)*, before considering whether the findings make it possible to determine if the sequence represents a collocation or a proper locution (see below). I go on to give an overview of the expression's distribution in time and space, and attempt to interpret its history from a variational viewpoint, that is considering its passages from one variety of French to another.

The verbal expression *tomber en amour* in linguistic and metalinguistic resources of French

What do French dictionaries say?

In general French dictionaries, *tomber en amour (avec quelqu'un)* appears as a “regionalism” in Quebec (TLF, Antidote) and in Canada (PR), and is given as being colloquial (“*fam.*”, PR, Antidote).

Two of the dictionaries also provide citations: TLF illustrates the use by quoting Tremblay (1982) (~ *avec quelqu'un*), and Antidote gives access to illustrations in online newspapers, *Québec Amérique* and *Le Soir*. The latter happens to be a Francophone Belgian journal; the context however reveals a reference to an Irish novel; the phrase's usage can hence be explained as a calque of Modern English; the choice of this citation to illustrate a Quebec usage does, however, seem inadequate and confusing.

According to the Quebec Lexical Index (ILQ), the expression is very largely attested in Canada (Acadia/New Brunswick, Quebec) and in the United States (Louisiana). It is mainly declined with the auxiliary *être* (rarely *avoir*), and can also be found in the pronominal form, *retomber en amour* (ILQ).

One also finds similar forms: *être en amour* as a ‘regional locution’ in Canada (PR) and in Quebec, more precisely in colloquial speech (Antidote) (Ø, i.e. absent from TLF). However, according to ILQ, this expression is less frequent in Canada than the variant with *tomber*.

Version d’auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

The expression does not appear in the other modern resources I used (nor does *être en amour*), whether it be in specialised dictionaries on fixed expressions (Lamiroy 2010, Bernet & Rézeau 2010a, 2010b) or in other dictionaries on diatopicisms (Ø DSR, DHFQ, DRF, DRFCompl, DRFA, DictBelg; also BDLP), with one notable exception: the *Dictionary of Louisiana French* (DLF). The latter explicitly describes both the sequence *en amour* “in love” (essentially corresponding to Modern French *amoureux*)² as well as *tomber en amour (avec)* “to fall in love (with)”, and illustrates its use with the auxiliary *avoir* (indeed dominant, if not exclusif in Cajun French), in absolute usage and with the preposition *avec*:

1. [Modern French (Louisiane)]
Et cette femme là alle amait mon oncle. Elle avait tombé en amour (DLF *amour*).
‘And this woman, she loved my uncle. She had fallen in love.’
2. [Modern French (Louisiane)]
Il a tombé en amour avec la musique cadien [orig.] (DLF *tomber 3*).
‘He fell in love with Cajun music.’

The existence of another locution, *être en amour pour même* “to be truly in love” (DLF *pour 5*), indicates both the vitality of the expression *en amour* in Louisiana and a certain lexical fixedness. As one can see, the phrase *tomber en amour* and its variant *être en amour* are not simply Quebecan, but more largely used in Modern French in North America.

Back in Europe, the *Französisches Etymologisches Wörterbuch* (FEW) does not confirm the expression’s presence in modern dialect glossaries, but gives Middle French *tomber en amour* by quoting “Est 1538-1552”. The cited author actually writes *entrer en amour* “in amorem incidere” in his French edition (1549: 32a), but indeed adds *tomber en amour* as a variant, in his Latin edition:

3. [Middle French] *entrer, ou tomber en amour* “incidere in amorem” (Estienne 1552: 667)

² This usage is of course distinct from the general phrase *en amour* “dans une situation d’amour” (cf. TLF *amour* IVA1b [loc.]), and does not refer to a sexual act either (cf. TLF IVA2 [en part.]).

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ?* What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

4. [presented as Middle French] *tomber en amour* “devenir amoureux” Est 1538-1552 (FEW XIII/1: 406a I.2 TUMB-)

Other Old and Middle French dictionaries do not register the expression *tomber en amour*, nor *être en amour*, nor its second element, *en amour* (Ø TL, DEAF, Cotgrave, DMF); the latter only gives *en amour* “de bon gré” (‘willing’) and “par faveur” (‘complaisant’) (DMF *amour* 1 A3a, A4a).

However, the *Glossaire des patois de la Suisse romande* (GPSR) gives the use of the adjective in the Suisse Valais canton:

5. [Suisse Roman dialect] adj. *è-namòçou*. *In-n amœ* “en chaleur” (V. Lourt), *i-n amour* (V Lens) || “S’applique par plaisant. à des personnes (V Miège [Valais canton])” GPSR 1924-1933 I: 367b *amour* I3

The expression’s use is here interpreted as a humoristic application to humans of a lexeme normally used to qualify animals in heat (but not taken over in GPSR 1977: 358a). The latter is indeed a very largely attested homonym in Middle and Modern French, and is also used in the Middle and Modern French sequence *être en amour* (FEW, DMF, Cotgrave *amour*)³.

The GPSR’s interpretation seems to indicate that Middle and/or Modern French *tomber en amour* is to be explained by metonymic evolutions from a lexical type referring to animals. Such transfers from the animal world to humans and vice versa are indeed plausible at any time, and not specifically French. In Middle English, for example, *love* is also used for “animal love”, and therewith *to be in love* “to be in heat” in reference to animals, attested since *a*1300 (*a*1250) (MED *l̥ve* 2c). It is, however, unlikely that such an evolution could explain the origin of the verbal expression *tomber en amour* in North America, where the obvious dominant factor seems to be language contact (see below).

What about its occurrence in Middle French? The verb *tomber* “to end up (in a certain state, with a result, ...)” is only known from 15th

³ Cf. *être en amour* “être en rut (des animaux)” (‘to be in heat (of animals)’ (since 1492), also with *entrer*; noun variant: *ameur* (since 15th c. norm.; mod. dialects: pic., frprov., Béarn) (FEW XXIV: 471b 5a-b AMÖR “liquid”; cf. DMF *amour* 2).

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

century texts onwards, first—and most largely—in a construction with + *en* + noun, like *desconfort*, *nécessité*, *maladie*, *malheur* etc. (from c.1429-1430 to 1489-1491, in eight texts, according to DMF *tomber* IID1 fig.).

However, the expression does not seem to replace an Old French **choir en amour* (Ø Gdf, TL, DMF).

What about the second part of the verbal phrase, *en amour*, used in reference to a person liking someone (in terms of attraction or friendship)? According to the *Dictionnaire du Moyen Français*, it actually appears in expressions with verbs such as *tenir*, *prendre* and *mettre* + pronoun + adverb *en amour* (since 14th-15th c., DMF *amour* 1 B3a, A2d, C1b fig.). The dictionary's "search in texts" function also allows retracing the phrase *avoir qqn en amour* "aimer, apprécier qqn" (c.1465-1470, DMF).

The lexicalisation of the above expression *en amour* seems to have taken place well before the 14th century, in the light of the existence of the verb *enamourer* from Old to Modern French (since c. 1180, FEW XXIV: 407b AMÖR), also in pronominal usage *s'enamourer*, equivalent to our *tomber en amour* (since 13th c., FEW), and early passed on to Middle English, where it is chiefly used passively, *to be enamoured* "to be in love" (since 1303, OED3 *enamour* 1; Ø AND)⁴.

As far as the Modern French period is concerned, in the light of the phrase's apparent absence in Europe except for the uncertain attestation in Belgium (but Ø VALIBEL), it is worthwhile considering the general modern equivalent *tomber amoureux*. The latter is dated from 1696 (FEW XIII/1: 406a), and seems to be an innovation of early Modern French (Ø DMF), analogue to constructions such as *tomber malade* (since 1489-1491, DMF *tomber* IID1 fig.). It enters into a well-established paradigm with phrases like *être amoureux* and *devenir amoureux*, already used in Middle French (c. 1361, DMF *amoureux* 1 IB1), and even found in Old French: *estre amorous* "aimer (qqn) d'amour" (end 12th c., cf. TLF < GdfC).

As far as the modern usage in North America is concerned, with regard to the widespread hypotheses of a calque on Modern English *to fall in love*, it is worthwhile looking into the latter's origin, in an *etimologia remota* perspective. The expression *to fall in love* (frequent with *with*) "to

⁴ Despite the TLF's etymological hypothesis for OF *enamourer* (*en-* + *amour* + *-er*)—other Romance languages showing equivalent formations (Span., Port. Prov. *enamorar*; Italian *innamorare*)—it may be worthwhile checking their history, and maybe considering a continuation of a panroman verb rather than an OF invention.

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

become enamoured” is only known from Middle English onwards, and attested since 1530 (Palsgrave [and not c. 1500]⁵, OED3 *love* P2b)—just before Estienne’s publication. It enters into a well-attested paradigm, with Middle and Modern English phrases largely obsolete nowadays, like *to be caught in love* and *to be taken in love*, as well as *to fall out of love (with)* and *to take love to* (variant of O.E. *to nim love to*) (OED3 P2c-d & P5). The latter is already rivalled in Middle English by its modern variant *to be in love with* (since 1577 OED3 [and not a1398, given by OED3 as the first attestation, though concerning a different sense])⁶. This is also the case of its equivalent of Romance origin (ME, ModE) *to be enamoured* (see above). Without giving a clear hypothesis on the expression’s history, the English reference dictionary simply invites the reader to compare with “Middle French *tomber en amour* (since 1538)” [that is: Estienne] (OED3 P2b)⁷.

For the origin of the English expression *to fall in love*, one can imagine a literal translation of some Middle French or Medieval Latin, but I know of no instances of manuscripts copying process where the English form clearly represents a calque of another language. In the light of the vitality of the above-cited verbs in Middle and early Modern English, *to fall in love with* seems to result from an internal independent process in Middle English, maybe analogue to its variant *to be in love with* (ME, mainly ModE, though attested later than the former). The data I queried do not confirm my original hypothesis of a calque on Middle French, nor do they exclude the influence of Modern English on Modern French on the new continent, and maybe in Europe.

What about the origin of Modern French *tomber en amour*? The above gathered data are far too scarce to allow any founded conclusion.

⁵ The first occurrence given by OED for this expression—ca 1500 (1437)—actually illustrates another expression: *to fall into love’s dance* (*Kingis Ouair*: xlv: “So ferr ifallyng into lufis dance”, OED3 *love* P2b), Middle English (ME) *love daunce* meaning “a dance for lovers” (cf. MED *l’ve* 4b).

⁶ OED3 *love* n.1 IIIIP2a [phrases] *in love (with)* “enamoured (of), filled with love (for)”, third example; the first example (a1398) illustrates the homonymous phrase (ModE *to be in heat*) in reference to animals.

⁷ It therewith corrects its former edition, which referred to “fr. *tomber en amour*, 15th c. in Littré” (OED2: 53c *love* VII.f)—where the reader only finds *tomber amoureux* “devenir amoureux” (Littré *tomber* 14); the original information “15th c.” remains mysterious.

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

Theoretically speaking, it could be an innovation of Middle French that did not attain particular vitality, and was therefore evinced by its general Middle/Modern French equivalent *tomber amoureux*. It may be a variant of Middle French *entrer en amour*, or a calque of the Latin equivalent, *incidere in amorem*; it may finally also have been directly created within the French language itself, as it integrates a well-established paradigm.

However, the information given by Estienne on a Middle French *tomber en amour*, taken over by FEW, could also represent a pure lexicographic artefact. It is absolutely necessary to access more textual data to verify how largely the phrase *tomber en amour* and sequences like *entrer en amour* were actually used in Middle and early Modern French in Europe, and when (and where) they were first used in the New World.

The expression in Francophone text collections

In order to verify the existence of the phrase *tomber en amour* in Middle and Modern French, the first important corpus to check is FRANTEXT. The latter essentially reveals modern attestations from the end of the 20th century, in France, but always in writings of authors in contact with Modern English: a writer from the South-West who has spent several years in Canada (Navarre 1981, 1986) and a well-mediatised Parisian writer (Forlani 1989). Another sequence actually appears in a translation of an English novel, due to a writer from New York (Mathews 2005)—who either translates himself, or is translated by his French wife (from the Lyon area).

Occurrences in translations from Modern English into Modern French, which I excluded from my analysis, can indeed be found in several text collections, whether they give access to the Web (FRWAC), to literary texts (like Shakespeare, translated by the Norman Le Tourneur 1782, Google Books), or to English speeches in the Canadian Parliament (transcribed in HANSARD). Such cases illustrate the influence of the Anglophone language and culture, either directly or indirectly. Modern French writers can indeed be influenced by Francophone Canadian writers and want to use Quebec French expressions to provide a stylistic semantically transparent reference to Canadian culture.

Apart from the above-cited occurrences of *tomber en amour*, FRANTEXT also provides one attestation before the 20th century, in the writings of a well-known philosopher from the South-West of France:

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

6. [Modern French] Se complaire outre mesure de ce qu'on est, en tomber en amour de soy indiscrete, est, à mon advis, la substance de ce vice (Montaigne 1592, *Essais* I, FRANTEXT).
'To be excessively pleased with what one is, so as to fall in indiscrete self-love, is in my opinion the substance of this vice.'

With this attestation, this time in discourse (and not meta-discourse), one knows about the sequence's availability for 15th-16th centuries French speakers. It is here used with a preposition *de* and a pronoun *soy* which specifies the object of love. However, it is also followed by an adjective modifying the noun *amour*, and its lexical fixedness is still not certain.

It is worthwhile looking into other corpora. The expression is very well attested in Canada from the second part of the 19th century onwards (1862-2007, FLI; but Ø QUÉBÉTEXT), and used in different discourse genres, such as in newspapers (EUREKA), at Parliament (HANSARD: transcribing spoken discourse in French [2003]), or in informal conversations, in the recent Quebec oral corpus CFPQ (2009).

It is one of the traditional literary corpora, conceived for the study of diatopicisms, which provides an occurrence of the expression *tomber en amour* for Switzerland—where the GPSR already signalled the use of the adjective *en amour* (see above, example 5). It occurs in a poetic novel where it is assigned to the narrator's speech, in a repetition aiming at attracting the reader's attention. The context reveals the polysemy of the verb *tomber*, used both in its literal and figurative sense:

7. [Modern French (Neuchâtel)] Et pour moi qui suis votre fiancée d'hiver, que faites-vous donc? / Moi qui tombe tombe tombe en amour, vos bras n'amortissent même pas ma chute! Vous me voyez tomber et ne faites aucun geste... (A.-L. Grobéty 1984, *La Fiancée d'Hiver*, edit 24 Heures: Lausanne: 324, SUISTEXT).
'And for me, who I am your wintrily fiancée, what are you doing after all? / Me, who am falling, falling, falling in love, your arms do not even absorb my fall! You can see me fall and do not react at all...'

The novel's author, from Neuchâtel where she has spent most of her life (1949-2010), has no particular bond to the English culture or to North America—nor can one find contextual reference to this area.

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

Recent corpora of journalistic discourse also afford surprises: the North-Eastern press corpus *Est Républicain* reveals two occurrences in passages relating the life of French people (a writer and a Lorrain farmer), and EUREKA provides an attestation in Lebanon's first francophone daily newspaper, *L'Orient-Le Jour* (2011); however, the latter does not allow identifying the author's identity. The two European attestations illustrate formally interesting usages:

8. [Modern French (Lorraine), within quotation marks] Toulinois d'origine, hier exploitant la ferme de la Vaux-Marie aujourd'hui à demeure au village où il a restauré avec bonheur une ferme datant de 1835 et 'tombé en amour pour les vieilles pierres', qui se proposa d'en effectuer bénévolement la remise en état [...] (*Est Républicain* 2002).
'Originally from Toul [in Lorraine], recently farmer at Vaux-Marie [a farm near Verdun, known for having been targeted in World War I], nowadays settled in a village where he volunteered to restore with pleasure a little farm dating from 1835, and 'fell in love with old stones [...].'
9. [Modern French (Lorraine?)] A l'instar de Kathy Moreau [author of *Poupées et Baigneurs Français en Celluloïd*, 1994], tombée en amour pour les poupées il y a une poignées [*sic*] d'années [...] (*Est Républicain* 2002).
'Just as Kathy Moreau [author of *Poupées et Baigneurs Français en Celluloïd*, 1994], fallen in love with dolls a few years ago [...].'

In other cases, however, the speakers' identities cannot be identified, or contextual information reveals explicit references to North America (*Est Républicain*, FRWAC 3 times). FRWAC also contains the expression in a passage where a speaker relates a stay in Toulouse; this does not however allow drawing any conclusions on the expression's areal distribution.

For the intermediate time period between the 16th and the 19th centuries, the open and heterogeneous text collections Google Books and GALLICA allow searching texts from specific time spans, despite their restriction to searching exact sequences only. Choosing texts anterior to 1801 in Google Books gives access to Montaigne's writing, to a translation, and to a usage in reference to animals (see above). However, it also provides an intermediate occurrence of *tomber en amour* in reference to humans, in a theatre play due to two French authors (one from Brittany, one from Paris), from the first part of the 18th century:

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

10. [Modern French] ARLEQUIN. / Tomber amoureux. Oh! pour celui-là, je ne l'avois pas encore entendu. / Mademoiselle RAFFINOT. / Hé! oui, tomber amoureux. Ne dit-on pas tomber malade? Or, comme l'amour est une maladie, on doit dire tomber amoureux &, tomber en amour, comme tomber en apoplexie. / ARLEQUIN. / Laissons-là le terme, & revenons à Dorimon (Le Sage/D'Orneval 1721-1734, *Le théâtre de la Foire ou L'opéra comique, avec Figures*, vol. III [1725?], chap. "Les Amours Déguisés", scene X: 444, in: *Œuvres choisies de Le Sage, avec Figures*, vol. XIV, Amsterdam/Paris, 1783).
'ARLEQUIN. Falling in love. Ah! as to him, I had not heard that yet. / Miss RAFFINOT. / Well! yes, to fall in love. Does one not say to fall ill? Yet, since love is an illness, one has to say to become enamoured &, to fall in love, like to fall in apoplexy. / ARLEQUIN. Let's leave the term there, & and get back to Dorimon.'

The second speaker in this passage discusses the syntagmatic usage of *tomber amoureux* with a complex metalinguistic comment, suggesting that *tomber en amour* may actually be more precise or more expressive. This citation may simply illustrate the authors' individual creativity, but can also indicate the coexistence of *tomber amoureux* and *tomber en amour* in French in Europe in the 17th-18th centuries.

If the above-presented data inform on the expression *tomber en amour*, the corpora also reveal the use of similar phrases like *être en amour* and its element *en amour* (as defined above). In Europe, the latter are actually more largely attested than the sequence *tomber en amour*. Amongst the traditional corpora, FRANTEXT illustrates the use of *en amour* from the first part of the 15th century onwards (Charles d'Orléans c. 1415-57)⁸, and provides fourteen occurrences of the expression *être en amour* in 19th-20th centuries texts. These are due to three writers from the West and South of France—mainly Alphonse de Chateaubriand (from Brittany) [1848], but also Claude Roy (who has grown up in his mothers region, the Charente) [1969] and Jean-Paul Chabrol (Languedoc) [1976]. Especially the first usages can hardly be explained by language contact with Modern English

⁸ Other sequences such as *être en amour avec qn* in Louise Labbé's writings (1555, FRANTEXT) do not have the same meaning (here equivalent to general Modern French *faire l'amour*).

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

to be in love, nor by some influence of North American on European writers wanting to create intertextual coherence or adopt some “Quebec style”, which is indeed the case of other usages illustrated in FRANTEXT (in Mathews’ and Forlani’s writings).

Verbal expressions like *être en amour (avec quelqu’un)* are of course very well attested in contemporary North America (QUÉBETEXT 3 times [1893], EUREKA, HANSARD 1995-2010), like the expression *en amour* in adjectival and absolute usage, also in formal discourse like at the Parliament (HANSARD). But when do these expressions first appear in the New World?

A recent corpus conceived in Ottawa (MCVF) shows that the postposed phrase *en amour* used in reference to a person liking someone (in terms of attraction or friendship) is actually attested in Old French, from the beginning of the 13th century onwards. It first occurs with the verb *tenir* + pronoun: *tenir quelqu’un/quelque chose en amour*, since Joinville’s diary (1225 [1305], Ouest/Champagne [North-East of France] in the so-called Manuscript A from Bruxelles: 370–antedating DMF, see above). This 13th century attestation is followed by Froissart’s chronicles, where the verbal construction occurs several times (1333 [1399]: 190, 472, 488). The same author also provides the first attestation of the sequence’s usage with the auxiliary *être*, in *être en amour avec qqn*:

11. [Old French (Pic.)] “[...] que il estoit trop françois et que il ne savoit estre en paix et en amour avoecques ses gens” (Jean Froissart (1333) [1399], *Chroniques, Premier livre*: 180, MCVF)

In North America, MCVF reveals first isolated attestations of the expression *en amour* in Marie Morin’s prose (1649 [1697], *Annales de l’Hôtel-Dieu de Montréal*); they do not however illustrate the same sense.

***Tomber en amour*: locution or collocation?**

Having gathered rich data from French dictionaries and text collections, from the Middle Ages to the beginning of the 21st century, it is possible to describe the Modern French diatopicism *tomber en amour* beyond its

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ?* What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

simplified apposition to general Middle and Modern French *tomber amoureux*.

The text data show that the verbal phrase *tomber en amour* (Middle French [uncertain]; dialectal Modern French) “to start feeling physically or emotionally attracted by a person or an object” is semantically transparent, and does not occur in a particular figurative meaning. If it is classified by the *Nouveau Petit Robert* as a *locution*, this attribution is not coherent with TLF’s definition of the term:

12. “groupe de mots constituant un syntagme figé” (TLF *locution* B2):
“groupe de mots pris souvent dans une acception figurée” (B2a) / “[...] ayant dans la phrase la valeur grammaticale d’un mot unique” (B2b)
‘word group constituting a fixed syntagm’: ‘word group often taken in a figurative sense’ or ‘having the grammatical value of a single word in a sentence’

Indeed, if the sequence has a certain degree of lexical fixedness, this fixedness is limited: the expression is composed of a verb *tomber* “to end up in a certain physical or moral state” and a lexicalised phrase *en amour* (Old and Middle French, dialectal Modern French), prototypically used in reference to a person who is or seems attracted by another person, either physically or affectionately. The sequence (which would have the status of a *locution* according to TLF’s definition) is semantically and structurally close to its general Middle and Modern French counterpart *amoureux* (cf. TLF IA); however, it essentially fulfils an adverbial function (especially in Old and Middle French), and has not developed the same semantic derivations and grammatical usages as *amoureux* (the latter also functions as a noun, TLF II). It is therewith not completely equivalent.

According to the resources on Middle and Modern French, the verbal phrase *tomber en amour* also has little lexical fixedness, as the second element can alternate with *amoureux* and the former with *entrer* without involving semantic change. In Modern French, it is also fairly little fixed morphosyntactically: it is mainly declined with the auxiliary *être*, but also sometimes with *avoir*, in North America (ILQ), and more specifically in Louisiana (DLF). In addition, the object of love is not necessarily specified (absolute usage); If so, it is mainly introduced by the preposition *avec*, which can however alternate with *pour* and *de*. The entity therewith represents a semi-fixed expression, more precisely a *collocation*,

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

according to current terminology in French phraseology (e.g. Tutin 2010 for a synchronic approach), that is, essentially speaking, a lexically constrained lexical unit that has a specific meaning. I therewith take the term in its qualitative sense, and not in its statistical sense as Hausmann has shaped it (also Blumenthal, e.g. 2008: 22).

More precisely, the collocation is composed of a *base* (French *base*)—here *en amour*—and of a *collocate* (French *collocatif*): here the verb *tomber*. As it encloses another collocation, *en amour*, that is an adverbial and adjectival collocation with a nominal base (composed of a prepositional collocate *en* and a noun base *amour*), the verbal collocation constitutes a *cas d'enclassement de collocation* ('framing collocation'), according to Tutin's analysis of the sequence (2010: 62f).

This verbal collocation coexists with other semantically close but distinct collocations, themselves formed with the base *en amour* and other collocates—both verbs and auxiliaries: *être*, *retomber* and *rester* (which can also precede *amoureux*). However, text data do not show the possible combination of the verbs *se montrer* (cf. TLF *amoureux* I) and *devenir* with *en amour*, nor of the verb *entrer* with *amoureux* (well attested with *en amour*). The collocation *tomber en amour* is definitely not simply a diatopically marked equivalent of *tomber amoureux*.

Which hypothesis for the distribution of the collocation in time and space?

According to the findings presented here, the collocation *tomber en amour* meaning “to start feeling physically or emotionally attracted by a person or an object” is diatopically marked in Modern French, where it is characteristic of North America (mainly 19th-21st c.). It is however also used outside this area, mainly by speakers who have a bond with the English language or French in North America, though not exclusively so.

From a diachronic viewpoint, text data show that it is the collocation *en amour* used in reference to a being who feels love or profound friendship, which first appears in writing, in adverbial usage, from the early 13th century onwards (MCVF see above). The collocation is probably not a calque of Modern English *in love*, but a survival of Old French. It has however not achieved greater vitality in general Middle and Modern French, especially as an adjective, where the adj. *amoureux* dominates. It

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

is also used as an adjective with different collocates: *être en amour* is equally attested from Old French onwards (cf. MCVF; example 11, above), and mainly in Modern French, in the 19th-20th centuries: first of all in North America, but also occasionally in Europe, in diatopically marked texts due to authors from the West and the South of France (see above). In chronological terms, this collocation is followed by the collocation *entrer en amour*, attested in Middle French lexicography (cf. Estienne), and again in North America since the 19th century (QUÉBETEXT; EUREKA).

As far as the collocation *tomber en amour* is concerned, it appears in France in the 16th century (Montaigne, Estienne) and again in the early 18th century, in a theatre play (Google Books), before large usage in Canada from the 19th to the 21st centuries. In Europe, one also observes sporadic occurrences in the 20th-21st centuries. Putting aside those that are probably to be explained by English or Canadian influence, the latter are mainly localised in the Eastern part of Francophone Europe (SUISTEXT, *Est Républicain*; maybe FRWAC); for the isolated form in North Africa (EUREKA), we have no data on the speaker. It is interesting to observe that the corpora attest greater formal variation in the rare European usages—*tomber en amour (pour/de/avec qn/qc)*—than in North America (mainly *avec*), where one however also finds *en amour* with collocates such as *rester* and *retomber* (ILQ, QUÉBÉTEXT, DLF).

One cannot exclude that both the collocation *en amour* and *tomber en amour* have been used without discontinuing from the 15th-16th century to the present, and the whole paradigm could have been exported to Quebec in the 17th century. This is where the sequence *tomber en amour (avec qn/qc.)* has achieved great vitality as well as a high degree of lexical fixedness, at least since the early 19th century. However, its first usage appears surprisingly late in North America (19th c.). In addition, the corpora do not illustrate sufficient occurrences before the 20th century, in particular from the 16th to the 18th centuries, to confirm continuity between a 15th-16th century sequence *tomber en amour* and the modern lexicalised diatopicism. This could be due to lack of sufficient and adequately prepared text collections available for analysis, even though historical data on French are exceptionally rich, compared to other European languages. This non-continuity rather has to be explained by the sequence's rareness in common usage.

The collocation *tomber en amour* in North America (18th-21st century) seems to be a calque of Modern English *to fall in love*. This calque has

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

however been facilitated by the existence in Modern French of the collocation *en amour* (mainly in North America). Sporadic usages elsewhere (like in Europe) are to be explained by North American influence, but also sometimes by the creativity of the French language itself. Some rare occurrences of the sequence *tomber en amour* in Europe in the 16th, 18th and 20th-21st centuries result from the availability of the verb *tomber* in general Middle and Modern French, as well as of the collocation *en amour* – and therewith of the verbal collocation *tomber en amour* in dialectal Modern French, and maybe in Middle French.

What do the gathered data show from a methodological point of view?

Certain corpora provide diatopic text annotation (notably MCVF, PFC, RÉGION), but these remain exceptional in French linguistics. In many cases, the researcher cannot identify the speakers' dialectal characteristics with certainty—as in the case of its isolated usage in the French Lebanon revue (cf. EUREKA). In addition, most corpora do not allow visualizing large enough extracts in order to submit the texts to a serious philological exam, nor do they reveal sufficiently reliable information for localising sequences with certainty. This particularly concerns web-drawn corpora such as FRWAC and I-FR, but also corpora which have not been conceived for linguistic analysis, like EUREKA. Nevertheless, the linguist can sometimes espy useful indicators, and interpret them carefully.

Other corpora are not accessible (BELTEXT); still others do not contain any occurrence of the collocation at all; indeed, the largest existing oral corpus, VALIBEL (cf. Dister *et al.* 2009), includes no occurrence of the sequence *en amour* in its texts anterior to 2000 for Modern French in Belgium. If one searches PFC (cf. Durand & Lyche 2003)—the only oral corpus giving access to dialectally annotated transcriptions of Modern French in different parts of the Francophone world in a pan-francophone perspective—it contains no more than eight occurrences even of the general Modern French adjective *amoureux*, *-euse* (in Africa, America, and Europe). This illustrates that such oral corpora are still too small for the analysis of diatopically marked phrases. In addition, amongst the rare oral corpora that are freely accessible, which allow search of word sequences, and which give access to a certain mass of speech productions (some half a million word-occurrences minimum)—which I consulted systematically—only the Canadian corpus CFPQ confirms usage of our collocation *tomber en amour*. Other corpora however illustrate its use in very different

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

discourse genres, and in highly formal, literary texts, but also in local daily news or informal exchanges.

The linguistic and metalinguistic data can actually be completed by observation: the verbal collocation is sporadically used in spoken discourse in the Eastern part of Francophone Europe, by speakers using it without any particular stylistic research, and having no particular bond to the Anglophone or North American culture. The expression has indeed been heard at the beginning of the 21st century in Switzerland (university student, Neuchâtel), and in informal situations in the South-East of France (elderly people and a pensioned teacher, Grenoble, 2011). Such empirical observation confirms the collocation's availability to speakers not only in North America, but also in Europe.

Conclusion

For the analysis of the diatopically marked Modern French collocation *tomber en amour*, existing dictionaries and corpora are numerous. However, documentation used in French lexicography has not traditionally been conceived for the study of diatopicisms, nor of more or less fixed expressions. One therewith disposes of little meta-linguistic data adapted to the study of French verbal expressions, whether it be properly fixed locutions or collocations.

The collocation *tomber en amour* is indeed hardly represented in modern French lexicography, even though it is its role to describe the language's entities with their syntagmatic and diatopic characteristics. Its scarce lexicographic representation seems to be partly due to different priorities of most French lexicographers and philologists.

Indeed, traditionally, the general aim of most editors of Old and Middle French texts is to give access to the manuscripts and copies they analyse. For this, semantically transparent sequences like the collocate *en amour* (Old and Middle French; dialectal Modern French) and the verbal expression *tomber en amour* (dialectal Modern French; Middle French uncertain) do not pose any particular problem. Their status as collocations also explains their absence from dictionaries on locutions.

The case study of the collocation and of its distribution in space and time points out the weaknesses not only of contemporary French dictionaries, but also of the heterogeneous text data that are available. It

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

also reveals in my opinion that French linguistics still lack a large corpus comparable to the *Oxford English Corpus* or to the *Corpus de Referencia del Español Actual*. Such a corpus would have to be annotated and balanced according to different pertinent criteria, notably diatopic, chronological, discursive and thematic, and would without doubt provide more reliable data for the synchronic and diachronic analysis of diatopically marked lexemes and expressions.

In the absence of such a corpus, the present work can still be completed by consulting the World Wide Web, which also gives access to large data on the French language in the southern *Francophonie*. For this area, to take an example, I checked the existing resources for a Francophone area in Central America, the Antilles—but found no certain data, neither in the only existing text corpus (ZOBEL), nor in the numerous Creole and French dictionaries (Wissner 2012b).

It would be necessary to check whether occurrences in the southern Francophone world only appear in the discourse of speakers where its usage can be explained by English influence (see Wissner 2012b for the Antilles), or whether one can advance endogenic usage. However, it is not always possible to realise critical philological text analysis with Google; in addition, this corpus does not provide a lemmatiser; and the linguist searching sequences of a certain frequency is overloaded by a mass of non pertinent data. Nevertheless, using such complementary corpora would certainly add decisive data to confirm or reject my hypotheses.

Bibliography

AND: *Anglo Norman Dictionary* [1977-1992], prepared in Aberystwyth under the supervision of D. Trotter; free online access, <http://www.anglo-norman.net/> (accessed 30th September 2012)

Antidote: Druide (collective) 2012, *Correcteur, Dictionnaires, Guides* [1996], software version 6.1, copyright 2012, Montreal (Quebec): Druide Informatique Inc.

Baroni, M. & S. Bernardini & A. Ferraresi & E. Zanchetta 2009, “The WaCky Wide Web: a Collection of Very Large Linguistically Processed Web-Crawled Corpora”, *Language Resources and Evaluation* 43(2), 209-226

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

BDLP: *Base de données lexicographiques panfrancophone*. Software composed of twenty national databases, Laval (Quebec): AUF/TLFQ, free online access, <http://www.bdlp.org/> (accessed 30th March 2013)

BELTEXT. Textual database on the French language and literature in Belgium composed of written and oral corpora, conceived by Ch. Delcourt *et al.*; no public access

Bernet, Ch. & P. Rézeau 2010a, *C'est comme les cheveux d'Éléonore. Expressions du français quotidien*, Paris: Balland

Bernet, Ch. & P. Rézeau 2010b, *On va le dire comme ça: dictionnaire des expressions quotidiennes*, second edition [2008], Paris: Balland

Blumenthal, P. 2008, "Französische Kollokationen in Lexikografie und Forschung", *Lexicographica* 24, 21-38

Cappeau, P. & F. Gadet 2007, "L'exploitation sociolinguistique des grands corpus. Maître-mot et pierre philosophale", *Revue Française de Linguistique Appliquée* 12(1), 99-110

CFPQ: *Corpus de français parlé au Québec*: multimodal Corpus of contemporary oral French in Quebec prepared at Sherbrooke University under the supervision of G. Dostie (2006-2012), gathering audio-visual recordings from 2006 to 2011 (accessible on-site) and aligned transcriptions of some 38.2181 words (on 24-05-2012); free online access, <http://pages.usherbrooke.ca/cfpq/corpus.php> (accessed 25th September 2012)

COD: Soanes, C. & A. Stevenson (eds.) 2008, *Concise Oxford English Dictionary*, twelfth edition [1911 *The Concise Oxford Dictionary of Current English*], Oxford: OUP

Cotgrave, R. 1611, *A Dictionarie of the French and English tongues*, London

CREA. *Corpus de Referencia del Español Actual*: Textual database of Current Spanish conceived by the *Real Academia Española*, gathering some 160 million words (state on 2008) of different text types from all hispanic varieties in Europe and America, annotated according to genre, region and theme; free online access, <http://www.rae.es> (accessed 30th July 2012)

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

DHFQ: Poirier, Cl. (ed.) 1998, *Dictionnaire historique du français québécois. Monographies lexicographiques de québécismes*, Sainte-Foy (Quebec): Laval UP

DictBelg: Francard, M. & G. Geneviève & R. Wilmet & A. Wirth 2010, *Dictionnaire des belgicisms*, Bruxelles: De Boeck-Duculot

Dister, A. & M. Francard & Ph. Hambye & A.-C. Simon 2009, “Du corpus à la banque de données. Du son, des textes et des métadonnées. L'évolution de la banque de données textuelles orales VALIBEL (1989-2009)”, *Cahiers de Linguistique* 33(2), 113-129

DLF: Valdman, A. & K. J. Rottet (eds.) 2010, *Dictionary of Louisiana French: As Spoken in Cajun, Creole, and American Indian Communities*, Jackson: UP of Mississippi

DMF 2012: *Dictionnaire du Moyen Français 1330-1500*, third edition [DMF1 2003, DMF2 2007], dictionary completed by a corpus gathering some 1.8400 pages of Middle French texts accessible with a software allowing to “search in texts” (individual and two consecutive lemma), Nancy: ATILF-CNRS, free online access, <http://www.atilf.fr/dmf/> (accessed 30th November 2012)

DRF: Rézeau, P. (ed.) 2001, *Dictionnaire des régionalismes de France (DRF). Géographie et histoire d'un patrimoine linguistique*, Bruxelles: De Boeck-Duculot

DRFA: Rézeau, P. 2007, *Dictionnaire des régionalismes du français en Alsace*, Strasbourg: Strasbourg UP

DRFcompl: Rézeau, P. 2007, “Compléments au DRF”, In: *idem* (ed.), *Richesses du français et géographie linguistique*, vol. I Bruxelles: De Boeck-Duculot, 415-479

DSR: Thibault, A. 2004, *Dictionnaire suisse romand. Particularités lexicales du français contemporain*, second edition [1997], Geneva: Zoé

Durand, J. & C. Lyche 2003, “Le projet ‘Phonologie du Français Contemporain’ (PFC) et sa méthodologie”, In É. Delais & J. Durand (eds.), *Corpus et variation en phonologie du français: méthodes et analyses*, Toulouse: Mirail UP, 212-276

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

Estienne, R. 1552, *Dictionarium latinogallicum*, third edition, Paris: R. Estienne, free online access, <http://www.cnrtl.fr/> (accessed 1st July 2011)

Estienne, R. 1549, *Dictionnaire françois-latin*, second edition [1539], Paris: R. Estienne

Est Républicain. Corpus consisting of articles of all regional and local editions of the French daily newspaper *L'Est Républicain* (North-Est of France) from 1999, 2002 and 2003; free online access, <http://www.cnrtl.fr/corpus/estrepublikain/>, used via the CQPWeb search engine at ATILF-CNRS, Nancy (accessed 30th September 2012)

EUREKA. Multilingual media corpus (formerly *Biblio Branchée*), gathering 6.023 referenced French sources covering local, national and international news (state on 06-01-2012); copyright 2012, access under licence, <http://www.eureka.cc/Default.aspx> (accessed 27th July 2012)

FEW: Wartburg, W. v. 1922-2002, *Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes*, 25 vol., Bonn: Klopp *et al.*

FLI: *Fichier lexical informatisé*: database for 16th to 20th centuries French in Quebec constituted of 400.000 files from 1975 to 1990 for the DHFQ project (Laval); free online access, <http://www.tlfq.ulaval.ca/fichier/> (accessed 27th July 2012)

Forlani, R. 1992, *Goultière* [1989], Paris: Gallimard

FRANTEXT. Textual database of some 248 million words from 4.084 referenced 16th-21st centuries French texts (state on 27-07-2012), belonging to the fields of sciences, arts, technology and most of all of literature, conceived for the TLF project (Nancy); online access under subscription, <http://www.frantext.fr/> (accessed 27th July 2012)

FRWAC. Mega French corpus with texts of various discourse genres (some 1.6 billion words) extracted from the web domain “.fr” by retrieval of word pairs within the WaCky Wide Web project (Trente/Bologne); free access on enquiry, <http://wacky.sslmit.unibo.it/doku.php?id=download>; used via the CQPWeb search engine (Hardy, Lancaster) at ATILF-CNRS, Nancy (accessed 27th July 2012)

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

GALLICA. The French National Library's digital library (Paris), gathering manuscripts scanned from 2007 onwards; free online access, <http://gallica.bnf.fr/> (accessed 27th July 2012)

Gdf, GdfC: Godefroy, F. 1880-1902, *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle*, 10 vol. (Gdf I-VII/1, GdfC VII/2-X), Paris: Viewieg/Bouillon

Google Books: Textual database scanned by Google, browsed 'before 1801' with limited preview and full view; free online access, <http://books.google.fr/> (accessed 26th September 2011)

GPFC: *Glossaire du parler français au Canada* 1930, Quebec: L'Action Sociale

GPSR 1924- : *Glossaire des patois de la Suisse romande*, currently published up to vol. 8 (*a-gonyon*), Paris: V. Attinger/Genève: Droz

HANSARD. French Corpus of the Canadian Parliament's House of Commons' Debates, constituted of the complete debates' proceedings (Chamber Business and plenary committee) from the 35th Parliament (Jan. 1994) to nowadays (41st Parliament), based on transcriptions of the deliberations' audio recordings and on reports by the service personnel (translated into French for originals in English); free online access, <http://www.parl.gc.ca/HouseChamberBusiness/ChamberSittings.aspx> (accessed 17th September 2011); non public integral texts of all debates, journals and the committees' procedures at the House of Commons from the 37th legislature's 1st session onwards (17-09-2001) checked on my behalf by the Parliament's publication services (open access scheduled)

I-FR. French corpus extracted from 50.000 pages of the Francophone Web in 2006 (no domain restriction) prepared in Leeds, gathering some 200 million words by retrieval of word pairs, filtered, lemmatised and automatically annotated with thematic clusters; search engine allowing retrieval of concordances and collocations; free online access, Corpus.Leeds.ac.uk/internet.html (accessed 25th July 2012)

ILQ. *Index lexicologique québécois* 2003. Inventory of French words in Quebec commented or studied from 1750 onwards, realised from 1979 to 1986 for the DHFQ project (Laval), under the supervision of Cl.

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

Poirier & L. Mercier; free online access, <http://www.tlfq.ulaval.ca/ilq/> (accessed 25th July 2012)

Lamiroy, B. (coord.) 2010, *Les expressions verbales figées de la francophonie. Belgique, France, Québec et Suisse*, Paris: Ophrys

Littré, É. 1872, *Le Littré. Dictionnaire de la langue française*, 4 vol. [1863-1872], suppl. 1877, Paris: Hachette

Mathews, H. 2005, *Ma vie dans la CIA : une chronique de l'année 1973*, Paris: P.O.L.

MCVF: *Modéliser le changement: les voies du français*: Corpus composed of mainly unabridged texts of different discourse genres and diatopic varieties from Old to Modern French, conceived under the direction of F. Martineau (Ottawa) (copyright 2009); free access on enquiry http://www.voies.uottawa.ca/corpus_pg_fr.html (accessed 27th November 2011)

MED: Kurath, H. & S.M. Kuhn (eds.) 1956-2002, *Middle English Dictionary*, 13 vol., Ann Arbor: University of Michigan Press, free online access (copyright 2001, updated 18th December 2001), <http://quod.lib.umich.edu/m/med/> (accessed 27th September 2011)

Navarre, Y. 1986, *Une vie de chat*, Paris: Albin Michel

Navarre, Y. 1981, *Biographie*, Paris: Flammarion

OEC: *Oxford English Corpus*. Textual database of the English language conceived for OED projects, gathering more than two billion words drawn from different genres, annotated according to text types, region and speaker's identities; copyright 2012, <http://oxforddictionaries.com/words/the-oxford-english-corpus> (accessed 30th November 2012)

OED3: Simpson, J.A. (ed.) 2000-2012, *OED Online*, third edition [1884-1928, *The Oxford English Dictionary*], Oxford: OUP, copyright 2012, online access under subscription, <http://www.oed.com> (accessed 27th July 2012)

OED2: Simpson, J.A. & E. Weiner 1989, *The Oxford English Dictionary*, second edition [1884-1928], 20 vol., Oxford: Clarendon Press

Parmentier, M. 2006, *Dictionnaire des expressions et tournures calquées sur l'anglais*, [Sainte-Foy, Quebec]: Laval UP

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

PF: *Phonologie du Français Contemporain: usages, variétés et structure*: Textual database conceived under the direction of J. Durand *et al.* (Paris); data gathered at 60 survey points in the Francophone world; some one million words transcribed available for query (state on 25-09-2011); free online access on enquiry, <http://www.projet-pfc.net>, copyright 2004-2008 (accessed 27th September 2011)

PR: Rey-Debove, J. & A. Rey (eds.) 2008, *Le Nouveau Petit Robert de la langue française 2009*, fourth edition [1967; 2007] Paris: Le Robert

RÉGION. Textual database of French Metropolitan regionalisms conceived for the DRF project (Nancy); regionally annotated corpus drawn from 220 works of 156 authors; accessible on-site, ATILF-CNRS, Nancy

Rézeau, P. 2007, "Des variétés dialectales gallo-romanes aux variétés régionales du français: la constitution d'un champ disciplinaire", D. Trotter (ed.), *Actes du XXIV^e Congrès International de Linguistique et de Philologie Romanes*, Tübingen: Niemeyer, 263-275

QUÉBÉTEXT. Textual database of contemporary French in Quebec constituted in the nineteen-nineties for the DHFQ project (Laval) from texts of various discourse genres from the 19th and 20th centuries (published from 1937, mainly literary); free online access to four corpora, www.tlfq.ulaval.ca/quebetext/ (accessed 27th July 2011)

SUISTEXT. Textual database of contemporary French literature containing all 129 works of fourteen 20th century writers from Switzerland, conceived under the direction of P. Knecht for the DSR project, accessible for members of the Centre of Dialectology and Regional French Studies at Neuchâtel University

TCOF. *Traitement de Corpus Oraux du Français*. Textual database of Metropolitan contemporary oral French corpora with aligned sound constituted from 2005 onwards in Nancy (mainly interviews, conversations and meetings); corpus gathering half a million word-occurrences, free online access (separate files), <http://www.cnrtl.fr/corpus/tcof/> (accessed 27th September 2011)

TL: Tobler, A. & E. Lommatzsch 1925-2002, *Altfranzösisches Wörterbuch*, 11 vol., Berlin: Weidmann, Wiesbaden: Steiner

Version d'auteur antérieure de : Wissner, Inka, 2014, « *Tombé en amour avec les corpus ? What Text Corpora Can Add to the Study of Diatopically Marked Expressions of Modern French* », in : Henry Tyne, Virginie André, Christophe Benzitoun, Alex Boulton et Yan Greub (edd.), *French through Corpora : Ecological and Data-driven Perspectives in French Language Studies*, Newcastle : Cambridge Scholars Publishing (également disponible comme livre électronique, par exemple via <http://www.ebrary.com/corp/>), 46-66, 21 pages.

TLF: Imbs, P. & B. Quemada 1971-1994, *Trésor de la langue française. Dictionnaire de la langue du XIX^e et XX^e siècle (1789-1960)*, free online access, <http://atilf.atilf.fr/tlf.ht2> (accessed 27th June 2011)

Tremblay, M. 1982, *Trois petits tours* [1971, Ottawa: Leméac], In Observatoire du français contemporain (ed), *Richesses et particularités de la langue écrite au Québec*, Montreal: Montreal University

Tutin, A. 2010, *Sens et combinatoire lexicale: de la langue au discours*, Grenoble, HDR thesis; free online access, w3.u-grenoble3.fr/lidilem/labo/file/HDR_Tutin.pdf (accessed 25th July 2011)

VALIBEL. Textual database of oral French in Belgium (Wallonia, Brussels) conceived at Louvain-la-Neuve (1989-2009), gathering more than four million words transcribed and aligned; public access to latest transcriptions (on enquiry under condition) currently suspended for administrative reasons; <http://www.uclouvain.be/valibel-corpus.html> (accessed on my behalf 2nd April 2013)

Vézina, R. 1998, review of DSR, *Langues et Linguistique* 24, 225-230

Wissner, I. 2012a, “Les grands corpus du français moderne: des outils pour étudier le lexique diatopiquement marqué?”, *SKY Journal of Linguistics* 25, 233-272; free online access, <http://www.linguistics.fi/julkaisut/sky2012.shtml> (accessed 26th March 2013)

Wissner, I. 2012b, “Le français à la Dominique dans le discours romanesque: reconstruction linguistique ou construction discursive?”, A. Thibault (ed.), *Le français dans les Antilles: études linguistiques*, Paris: L’Harmattan, 141-206

ZOBEL. Textual database of contemporary French literature in the Antilles composed of the Martinican writer Joseph Zobel’s prose works, conceived by A. Thibault (11 texts, 1946-2002: 2044pp.), constituted and annotated by P. Drouin; accessible on enquiry