

STUDI E SAGGI LINGUISTICI

LV (2) 2017

rivista fondata da

TRISTANO BOLELLI



Special Issue

WORD COMBINATIONS:
PHENOMENA, METHODS OF EXTRACTION, TOOLS

edited by

RAFFAELE SIMONE - VALENTINA PIUNNO

Edizioni ETS



Modelling French idioms in a lexical network

MARIE-SOPHIE PAUSÉ

ABSTRACT

This article proposes a model of description of idioms based on the Explanatory Combinatorial Lexicology's framework. It combines identification of idioms' lexical components and identification of dependency links between these components to form what we call a lexico-syntactic structure. This model will permit to predict formal variations of idioms thanks to the correlation of these variations with lexico-syntactic structures of idioms, but also with their lexicographic definitions.

KEYWORDS: idiom, lexicology, phraseology of the French language, Explanatory Combinatorial Lexicology.

1. *Introduction*

Idioms have been traditionally considered as prototypical phraseological units¹. Following Mel'čuk (2012: 37), we define them as semantically non-compositional phrases². As such, they are lexical units and, consequently, they must have their own entries in a lexical resource³. However, their status as phrases requires a specific description if we want to give speakers all the information they need to use them properly or to combine their constituents correctly. Nevertheless, we notice that lexicographic descriptions of French idioms are often disparate. For example, they frequently do not have their own lexicographic articles.

Nowadays, with the advent of modern information technology and processing, a new kind of lexicography has emerged. The *French Lexical Network* (fr-LN) (Lux-Pogodalla and Polguère, 2011) is being developed in ATILF⁴ laboratory (Nancy, France), in the line of resources like *FrameNet*

¹ «Idioms form the majority and may be regarded as the prototype of the phraseological unit» (GLÄSER, 1998: 126).

² A phrase is a combination of at least two lexical units linked by syntactical dependencies.

³ This idea is shared, among others, by BALLY (1909), KAVKA and ZYBERT (2004).

⁴ *Analyse et Traitement Informatique de la Langue Française*: <http://www.atilf.fr/>.

(Ruppenhofer *et al.*, 2010) or *WordNet* (Miller *et al.*, 1990). fr-LN's main feature is that it is worked out within the Meaning-Text Theory (MTT)'s framework. More in particular, it employs the Explanatory Combinatorial Lexicology's tools (ECL). But it also shares at least one feature with the other networks just mentioned: one of its aims is systematically describing idioms like any other lexical unit.

After introducing the features of idioms as lexical units (§ 2), we will briefly present our theoretical framework and the fr-LN's main principles (§ 3). We will, then, describe our modelling of idioms, which combines a precise identification of lexical constituents, with a clear-cut description of their syntactic structure (§ 4).

Lexical units (lexemes or idioms) will appear in small caps (e.g., TO CATCH THE EYE) and meanings will be written between single quotes (e.g., 'to attract attention'). Most of the examples used come from Frantext or FrWac⁵.

2. Features of idioms as lexical units

2.1. General features

As semantically non-compositional units, idioms, like any other lexical unit, must have a lexicographic definition. For example, NOYER LE POISSON in example (1) (lit. "to drown the fish"; Eng. TO CLOUD THE ISSUE, It. IMBROGLIARE LE CARTE) can be defined like this :

X noie le poisson: X, who has to express himself on something that is problematic, avoids speaking of the problem, by telling irrelevant things.

- (1) Il m'a même dit que c'était pas sa... cette femme, là, qui avait pris l'appel, que c'était un secrétariat, qu'on l'avait prévenu qu'à neuf heures et demie, enfin je sais pas ce qu'il m'a raconté, j'ai pas très bien compris, mais c'était pour *noyer le poisson*, sûrement... (Frantext)

"He even told me that it wasn't his... that woman, who had taken the call, that it was a secretariat, that he had been warned only at 9.30, after all I don't know what he told me, I didn't understand very well, but he wanted to *cloud the issue*, surely..."

⁵ *Frantext* is a text database that contains 4516 references: <http://www.frantext.fr>. *FrWac* is a web database that contains 1,613,206,614 tokens (BARONI *et al.*, 2009).

Furthermore, most idioms have some semantically equivalent lexis. *EMBROUILLER* ‘to confuse’, *TOURNER AUTOUR DU POT* (lit. “to turn around the pot”; Eng. *TO BEAT ABOUT THE BUSH*, It. *PRENDERLA ALLA LARGA*) and *BOTTER EN TOUCHE* (lit. “to kick in touch”; Eng. *TO KICK INTO TOUCH*, it. *SALVARSI IN ANGOLO*) are quasi-synonyms of *NOYER LE POISSON*. Therefore, idioms convey paradigmatic links with other lexis. These links can also be syntagmatic, as is the case between *BOIRE* and *JUSQU’À PLUS SOIF* (lit. “until no thirst”) ‘in large amount’ in example (2).

- (2) Kosita est un homme complètement défiguré, brûlé à la guerre, qui *boit* de la vodka *jusqu’à plus soif*. (FrWac)
 “Kosita is a completely disfigured man, burned during the war, who drinks big quantities of vodka.”

In this instance the syntagmatic links are collocational, between a base (*boire*) and a collocate (*jusqu’à plus soif*) (Mel’čuk, 1998).

Finally, just like lexemes, idioms can be described in terms of polysemy and homonymy. For example, *COUP DE FEU* has at least four senses. But we can’t establish semantic links between all of them. The main definitional principle of polysemy is, then, not respected. In fact, we can identify two vocables⁶, as showed by Figure 1⁷.

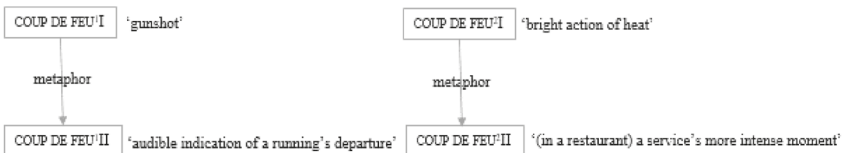


Figure 1. *Polysemous structure of COUP DE FEU¹ and COUP DE FEU².*

2.2. Specific features

Idioms’ specificities as lexical units reside in the fact that they are formally phrases. As a first consequence their grammatical features can’t be described like those of lexemes. The main grammatical feature attributed to all lexemes is a part of speech (PoS hereafter). PoS permit to predict a good

⁶ A vocable is a set of lexical units linked by polysemy.

⁷ For more information about our treatment of polysemy, especially that of idioms, see PAUSÉ and SIKORA (2016).

part of a lexeme's combinatorial proprieties – namely the ways how it can combine with other lexical units in a sentence. As a verbal phrase, NOYER LE POISSON takes the place of the predicate in the sentence. Consequently, the PoS of idioms will be the one of the governors of their phrases: NOYER LE POISSON is a verbal phrase, therefore a verbal idiom. By analogy, the PoS of LES DOIGTS DANS LE NEZ (lit. “fingers in the nose”) ‘easily’ is *nominal idiom*. Of course, we need to specify that it behaves like an adverb, as showed by example (3).

- (3) Ses cinq enfants MERVEILLEUX sont tous BRILLANTS, réussissent à leurs examens *les doigts dans le nez*, trouvent des jobs ÉPATANTS, ne se DISPUTENT JAMAIS avec leurs parents, etc. (Frantext)
 “His five WONDERFUL kids are all BRILLANTS, pass their exams *easily*, find SPENDID jobs, never argue with their parents, etc.”

This classification of idioms that do not behave like their phrase's head counters the traditional view of grammatical description. Most approaches tag LES DOIGTS DANS LE NEZ as an adverbial idiom. According to us, this type of grammatical characterisation based on the behaviour of idioms in sentences doesn't allow to describe idioms such as UN PEU ‘a few, some, a little’, which can behave like an adverb in (4a) or a determiner in (5a), but is formally a nominal phrase. This is proven by the fact that the adjective PETIT ‘little’ can modify its nominal component, like in examples (4b) and (5b).

- (4) a. Ses cinq enfants MERVEILLEUX sont tous BRILLANTS, réussissent à leurs examens *les doigts dans le nez*, trouvent des jobs ÉPATANTS, ne se DISPUTENT JAMAIS avec leurs parents, etc. (Frantext)
 “His five WONDERFUL kids are all BRILLANTS, pass their exams *easily*, find SPENDID jobs, never argue with their parents, etc.”
 b. Michel semble *un petit peu* trop occupé pour me rendre visite.
- (5) a. Ses cinq enfants MERVEILLEUX sont tous BRILLANTS, réussissent à leurs examens *les doigts dans le nez*, trouvent des jobs ÉPATANTS, ne se DISPUTENT JAMAIS avec leurs parents, etc. (Frantext)
 “His five WONDERFUL kids are all BRILLANTS, pass their exams *easily*, find SPENDID jobs, never argue with their parents, etc.”
 b. Elle a fait chauffer *un petit peu* de café.

Let us also remind that free nominal phrases can also play the role of an adverb, as in example 6.

- (6) Elle est repartie, *le sac à dos sur les épaules*.
 “She left, her backpack on her shoulders.”

The example of *UN PEU* highlights another typical feature of idioms: their variations are not, like lexemes, only morphologic, but also syntagmatic. Traditionally, variations of any type – paradigmatic or syntagmatic – were not admitted. Frozenness was considered as one of the main definitional criteria for idioms. This led linguists to drop out all forms of syntagmatic modification, especially as regards verbal idioms, as illustrated hereafter with *PRENDRE LA TANGENTE* (lit. “to take the tangent”) ‘to leave’ (example from Gross, 1996: 12).

- Pronominalization: **Paul l’a prise*. (lit. “Paul took it.”)
- Dislocation: **La tangente, Paul l’a prise*. (lit. “The tangent, Paul took it.”)
- Cleft constructions: **C’est la tangente que Paul a pris*. (lit. “It’s the tangent that Paul took.”)
- Relativization: **La tangente que Paul a prise*. (lit. “The tangent that Paul took.”)
- Modifications of a constituent other than head: **Paul a pris la longue tangente*. (lit. “Paul took the long tangent.”)

However, in the light of corpus linguistics, some noted that almost all idioms’ variations our linguistic intuition would exclude were substantiated by oral and even written productions (Burger, 1998: 23; see also Nunberg *et al.*, 1994; Abeillé, 1995 and Fellbaum, 2014). This confirms Moon’s (1998: 2) remark: «Fixed Expressions [...] is unsatisfactory as a term, since it will be seen that many fixed expressions [...] are not actually fixed».

Examples below show syntagmatic variations in some French verbal idioms, such as passivization in (7), dislocation in (8), cleft construction in (9) and modification of a non-head constituent in (10).

- (7) Mesdames, Messieurs, il y a quelques mois de cela, la révolution grondait pour empêcher le gouvernement de nous mettre des bâtons dans les roues en installant les fameux radars fixes. À l’aube de 2007, *le poisson a été noyé* et nous avons accepté le fait d’être fliqués par des flashes et des appareils photographiques automatiques. (Web)

“Ladies and Gentlemen, a few months ago, revolution erupted, to prevent the government from throwing some sand into our wheels by installing the famous fixed radars. On the eve of 2007, the issue was cloud (lit. the fish was drown) and we accepted to be tracked by flashes and automatic cameras.”

- (8) Kevin Régimbald a fait le saut au football professionnel. En début de saison, les Roughriders de la Saskatchewan ne l’avaient pas retenu dans la formation, mais lui avaient gardé une place dans l’équipe de réserve. Finalement, Régimbald a eu sa chance. *La glace, il l’a brisée* le 17 août dernier, et pour ajouter au stress d’un premier match dans la Ligue canadienne de football, c’est contre les Alouettes de Montréal qu’il a fait ses premiers pas. (Web)
 “Kevin Régimbald got into professional football. At the season’s beginning the Saskatchewan’s Roughriders hadn’t admitted him in the reserve team. Finally, Régimbald had an opportunity. (Lit.) *The ice, he broke it* on august the 17th, and in addition to the stress of a first march in the Canadian league of football, it’s against the Montreal Alouette that he took the first steps.”
- (9) Si on veut retrouver la confiance et le respect du peuple, *c’est devant sa porte qu’il faut commencer à balayer*. (Web)
 “If we want to recover people’s trust and respect, it’s in front of our door that we have to begin to sweep.”
- (10) Le Nouvel Obs ajoute que B., en fait, règle ses comptes, dans son livre : évincé par le tribunal qui n’a pas voulu le suivre dans sa plainte en diffamation contre S., il lui *taille un costume sur mesure* dans le livre. (Web, example anonymized)
 “Nouvel Obs (French newspaper) adds that B., in fact, settles scores, in his book: as he was ousted by the court that didn’t want to follow him with his libel suit against S., he (lit.) *cuts him a tailor-made suit*.”

In example (10), the modified idiom is *TAILLER UN COSTUME* ‘to criticize somebody’s actions openly in order to prejudice his reputation’. We can see that *SUR MESURE* ‘tailor-made’, a regular modifier of *COSTUME* ‘suit’, is used to intensify the whole meaning of the idiom: *tailler un costume sur mesure* means ‘to criticize [...] a lot’. The lexical meanings of constituents are then partly reactivated – we can speak of *remotivation*⁸ (Fr. *défigement*; Yakubovitch, 2015; Lecler, 2006) – but the idiom’s meaning is preserved. This shows that a complete lexicographic description of idioms must consider their syntagmatic aspect; in other words, to master the correct use of an idiom, a speaker must know which types of variations this idiom can undergo.

⁸ See § 4.1. for another example.

A syntagmatic description of any phrase is closely bound to its lexical constituents. Indeed, the valency of the lexis (that is to say their capacity of governing or being governed by other lexis) determines how they combine to form phrases.

To be able to generate all syntagmatic forms of an idiom, we must adopt a lexico-syntactic description. We will propose such a description after having presented our theoretical framework.

3. *Theoretical framework and French Lexical Network*

3.1. *Theoretical framework*

This work is based on the Meaning-Text Theory, and more particularly the Explanatory Combinatorial Lexicology's principals (Milićević, 2006).

The Meaning-Text Theory is a theoretical framework for the description of natural languages which allows to modelise the correspondance between meanings and texts. A Meaning-Text model starts with a semantic representation and ends out with a phonological representation.

The Explanatory Combinatorial Lexicology is the lexicographic part of the Meaning-Text Theory (Mel'čuk *et al.*, 1995). A model of lexicon based on the Explanatory Combinatorial Lexicology's principals allows a Meaning-Text Model, in conjunction with a formalized grammar of the language, to establish correspondances between a semantic representation and a syntactic representation.

The aim of our work is to build a model that permits to generate all syntagmatic forms of an idiom. This can only be done in a model of the lexicon that precisely describes lexical unit's combinatory. For our work, we use and expand the data of the French Lexical Network (hereafter, fr-LN).

3.2. *French Lexical Network*

The fr-LN is an implementation of a French lexicon's model⁹ worked out in this framework and modelled under the name of *lexical system* (Polguère, 2009; Polguère, 2014).

⁹ In parallel, similar resources are being built, for English and, in a less advanced state, for Italian, Korean, Russian and Spanish.

We will just introduce the main features of the fr-LN. This network forms a graph, made of nodes mainly occupied by lexical units, or *lexis* – lexemes and idioms – and linked by arcs representing lexical relations of different types: paradigmatic links, like synonymy, or syntagmatic ones, like collocations, for example. The graph’s structure is constantly evolving, according to the lexicographers’ work, which allows, *inter alia*, to add new nodes and lexical links between lexis. In the end, each node will be connected to at least one other node. A network tends to represent, as noted by Kiss (1968), a speaker’s mental lexicon.

Nodes include a lexicographical description that encompasses:

- grammatical properties (part of speech, genre, usage notes, formal characteristics, syntactical position, etc.);
- morphological templates (Gader *et al.*, 2014);
- government patterns;
- definition – in development, in accordance with principals described, *inter alia*, in Mel’čuk and Polguère (2016);
- examples;
- paradigmatic and syntagmatic links with other lexical units (modelled with lexical functions; see Mel’čuk, 1998).

Some of these items are illustrated by Figure 2.

<ul style="list-style-type: none"> ○ ogarette ● ogarette I ● ogarette II ● ogarette III 	<p>[CG]</p> <p>nom commun fém</p> <hr/> <p>[DF]</p> <p>chose fabriquée pour être fumée</p> <p>cigarette destinée à être fumée par $X_{c,1}$</p> <p>=</p> <p>objet t_1 destiné à être fumé III_a par l’individu X</p> <p>de forme allongée</p> <p>constitué d’un petit tube de papier très fin rempli de tabac^{1,2} haché, que x fait se consumer par une de ses extrémités</p> <p>que x tient entre ses lèvres $I,1$ ou entre ses doigts $I,1_a$ quand x fume</p> <hr/> <p>[FL]</p> <p>Syn : vieille fam cibiche, argot fam garo, argot fam clope, argot fam tige III, rég fam schmer, vieillissant fam sèche, vieillissant fam cancéreuse II</p> <p>Syn₅ : roulée II; blonde II, brune II</p> <p>Contr : cigare I; pipe I_a</p> <p>S_{1c}^{usual} : fumeur, fumeuse</p> <p>S_{inst}^{Real,1} : allumette $I,1$; briquet; allume-cigares</p> <p>S_{inst}^{Real,II} : fume-cigarette</p> <p>S_{loc} : bureau de tabac, tabac¹ III, débit de tabac</p> <p>Mult : paquet $1c$ [de -s] < cartouche III,b [de -s]</p> <p>Caus,Func₀ : rouler¹ III,2 [ART -]</p> <p>S₁^{usual}CausFunc₀ : cigarettier a</p>
--	---

Figure 2. Sample of article-view of CIGARETTE I in the fr-LN; grammatical characteristics (CG), definition (DF), lexical functions (FL).

Idioms receive the same type of description, as showed by Figure 3.

<p>noyer le poisson</p>	<p>[GC]</p> <p>locution verbale [noyer_v I + le_{Art} + poisson I.a] locution forte</p> <hr/> <p>[DF]</p> <p>avoir un certain comportement</p> <p>X_n, noie le poisson</p> <p>=</p> <p>x devant s'exprimer à propos de Alpha qui pose problème</p> <p>x évite de traiter du problème lié à Alpha en disant des choses non pertinentes</p> <hr/> <p>[LF]</p> <p>Syn_n : botter en touche, tourner autour du pot, embrouiller, louvoyer II, tergiverser ; amuser le tapis</p>
-------------------------	--

Figure 3. Sample of article-view of *NOYER LE POISSON* in the fr-LN; grammatical characteristics (CG), definition (DF), lexical functions (FL).

The fr-LN is not the only lexical network that takes in account idioms as lexical units. Idioms are also well considered in FrameNet (Ruppenhofer *et al.*, 2010) and WordNet (Miller *et al.*, 1990).

FrameNet is a lexicographical resource for English and other natural languages, based on Construction Grammar (Fillmore, 1988) and Frame Semantics (Fillmore, 2008). Each frame defines a situation and its participants. For example, *SPILL THE BEANS* 'to divulgate a secret' implies three participants: a speaker, an interlocutor, and an information revealed. This information has to do with a subject, and is communicated by oral or written channel. In FrameNet, each lexical unit like *SPILL THE BEANS* is described according to the constructions in which it occurs.

This principal entertains the common image that each lexical item carries with it instructions on how it fits into a larger semantic-syntactic structure, or, alternatively, on how semantic-syntactic structures are to be built around it. (Fillmore, 2008: 49)

WordNet is also built for different natural languages, including English. It consists in an accurate description of synonymy. Idioms and lexemes with same meaning are classified in *synsets*, from the more neutral to the more stylistically or pragmatically marked.

FrameNet and WordNet are both principally oriented on semantics, even if FrameNet is at interface between semantics and syntax. The fr-LN offers a description of lexis' combinatorial proprieties (paradigmatic and syntagmatic links, and syntactic construction) compatible with a linguistic model of synthesis as the Meaning-Text Model. This model can be exploited

in order to predict the formal variations of idioms (Pausé, in preparation). In addition to a semantic description, idioms must receive a lexico-syntactic description.

4. *Lexico-syntactic structures of idioms*

A lexico-syntactic structure (hereafter LSS) is an association between several lexical units and a linear syntactic pattern. In other words, associating a LSS to an idiom amounts to speculating on the free phrase the idiom is built on. For example, *NOYER LE POISSON* comes from a fishing technique metaphor referring to exhausting a hooked fish by putting its head alternatively in and out of water¹⁰ (so as to drown it). Then, the LSS of *NOYER LE POISSON* is formed by the lexis *NOYER* ‘to hold someone under water in order to prevent him from breathing’, *LE* ‘the’ and *FISH* ‘animal provided with fins that lives in water’. Each linear syntactic pattern corresponds to a complete syntactic structure, i.e. parts of speech connected by functional links. This corresponds to a dependency grammar’s representation.¹¹ The main principle of dependency grammar is that dependency structures are based on the valency of lexical units. We can speak of *active* and *passive* valencies (Iordanskaja and Mel’čuk, 2009: 151). The former valency is composed of all actants that the lexis can control. The latter corresponds to all actants that can control this lexis. Figure 4 shows the assignment of a LSS to *NOYER LE POISSON*.

We can see that a LSS identifies precisely which lexeme was used to create the idiom. So it gives an access to lexicographical information of components lexemes: definition, grammatical characteristics and combinatorial indications.

4.1. *Lexical units’ identification*

The main problem in identifying lexical constituents is that it isn’t always easy to determine the origins of an idiom. Sometimes, specialised books provide no specific knowledge or give diverse explanations. In other cases, the historical explanation is in competition with a reconstituted origin

¹⁰ This explanation comes from REY and CHANTREAU (2007: 648).

¹¹ Dependency grammar’s principals are exposed (and opposed to constituency analysis), *inter alia*, in GERDES and KAHANE (2013) (see also KAHANE, 2012).

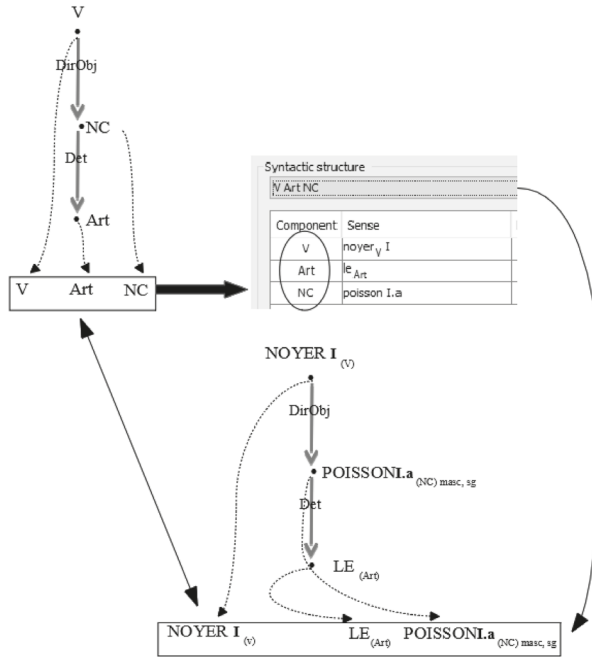


Figure 4. Assignment of a LSS to NOYER LE POISSON.

made by speakers of a language. For example, in French, we can say about someone who isn't in good shape that he isn't *dans son assiette*. If we try to translate it in English literally, we are immediately confronted to an etymological problem: where does the idiom come from? Is it a metaphor based on a boat's trim, which is called in French *l'assiette du bateau*? Or is it based on a vehicle (especially a plane)'s attitude (*l'assiette d'un avion*)? Do we rather have to take a look at horse riding, where we speak of a rider's seat (*l'assiette du cavalier*)? According to Rey and Chantreau (2007), the origin of this idiom has to do with an old sense of ASSIETTE, illustrated in example (10), that denotes someone's physical or moral disposition. We would translate it in English by MOOD, even if this lexis is less accurate.

- (10) [J]e crois que je ne ferai pas mal de me retirer; je sens que je ne suis pas ici dans mon *assiette* ordinaire. (Beaumarchais, *Le Barbier de Séville*, III, 11; quoted by Rey and Chantreau, 2007: 36)
 "I think I may leave; I feel that I'm not in my ordinary mood."

But if we have a look at the speakers' uses of the idiom, in particular at their remotivations, we realize that they activate another sense of ASSIETTE equivalent to Eng. PLATE:

- (11) a. Je n'étais pas *dans mon assiette*. *Elle est profonde, mon assiette, une assiette à soupe*, et il est rare que je n'y sois pas. (Frantext)
 "I wasn't (lit.) *in my plate*. *My plate is deep, a soup plate*, and it's rare that I'm not inside."
- b. Mardi 10 octobre à 19 h 30, la fondation PiLeJe organise une conférence-débat sur les bonnes habitudes alimentaires pour le bien-être et la santé, intitulée *Je me sens bien dans mon assiette*. (FrWac)
 "On October the 10th, at 7.30 pm, the PiLeJe foundation is organising a conference-discussion about good eating habits for wellness and health, entitled *I feel good* (lit.) *in my plate*."

Examples in (11) show that folk etymology (the reconstitution of a lexis's origin by the speakers) links DANS SON ASSIETTE to a metaphor based on alimentary balance, more than physical and moral one.

The lexicographer's problem is then to choose which lexical unit will be identified in the LSS: ASSIETTE 'mood' or ASSIETTE 'plate'? In our opinion, a complete lexicographic description may ally diachronic and synchronic views. Thus, we should identify both senses. For the moment, only synchronic view is taken in account, in order to predict remotivation.

Otherwise, we noticed that some idioms contained other idioms. For example, NE PAS LEVER LE PETIT DOIGT in example (12) (lit. "not to raise the little finger") 'not to do absolutely anything' encloses the idiom PETIT DOIGT (Eng. LITTLE FINGER, It. MIGNOLO).

- (12) J'ai bien senti que mon père *n'aurait pas levé le petit doigt* si ce commissaire avait exécuté sa menace et m'avait envoyé au Dépôt. (Frantext)
 "I felt sure that my father wouldn't have done anything if this commissioner had executed his threat and had sent me to the police station."

This fact implies that in some cases we have to identify the idioms' constituents indirectly: NE PAS LEVER LE PETIT DOIGT is composed of NE, PAS, LEVER and PETIT DOIGT, the latter being composed in turn of PETIT and DOIGT.

LEVER LE PETIT DOIGT also underlines another problem in identification of lexical constituents: how will we model French negation? NE and PAS function as a whole. But they also work as autonomous units, as showed by examples in (13).

- (13) a. Je suis allé jeter un œil dans la cuisine pour vérifier que les trucs collaient *pas* dans le fond de la casserole, mais tout se passait à merveille. (Frantext)
 “I had a look in the kitchen, to check that things didn’t stick to the pan, but everything was ok.”
- b. Je *ne* puis demeurer loin de toi plus longtemps. (Frantext)
 “I can’t stay far away from you anymore.”

NE and PAS, when occurring in combination, are modifiers of their verbal governor. Consequently, no direct syntactic relations link them. In order to reflect the associative link that any speaker makes between the two negative adverbs, we consider *ne...pas* as a *slot phraseme ne [X] pas*, where [X] is a verb. Slot phrasemes are close to the notion of *construction* as considered in construction grammar (Croft and Cruse, 2004).

4.2. Syntactic patterns’ identification

Syntactic patterns have two facets: the first consists of a linear succession of parts of speech that forms what we can call an elementary syntactic pattern. NOYER LE POISSON’s pattern is V ART NC (verb + article + common noun). The second facet will correspond to a complete syntactic structure, i.e. parts of speech connected by functional links. The active valency of NOYER is $X \sim Y$. Figure 5 illustrates the dependency structure involved.

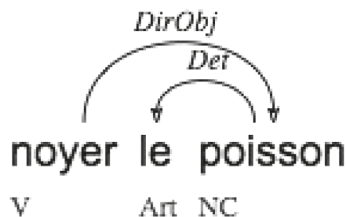


Figure 5. NOYER LE POISSON’s dependency structure.

Dependency links between units can predict a good part of their linear variations. For example, most transitive verbs can be passivized¹², namely, their object can turn into subject. Thus, a free phrase constructed on NOYER ‘to kill somebody by asphyxia, by immersing him in a liquid’ can be passivized (14a). The verb’s propriety is then transferred on the idiom NOYER LE POISSON (14b).

- (14) a. Ève apporta la bouteille renflée où une vipère *avait été noyée* dans l’eau-de-vie de prune, lâchant son venin pour renforcer l’alcool.
 “Eve brought the bulging bottle in which a viper had been drown in plume aqua vitae, releasing its venom to enhance alcohol.”
- b. À l’aube de 2007, le poisson *a été noyé* et nous avons accepté le fait d’être fliqués par des flashes et des appareils photographiques automatiques.
 “At the beginning of 2007, (lit.) *the fish was drown* and we accepted to be tracked by flashes and automatic cameras.”

With an accurate description of the lexical and syntactic proprieties of idioms, we will hopefully be able to predict a good part of their variations. The idea is to compare their lexico-syntactic structures to those of corresponding free phrases. We will then see whether idioms with the same lexico-syntactic structure accept the same variations or not, and, if the answer is negative, try to explain why.

For example, if we compare NOYER LE POISSON with SUCRER LES FRAISES (15a) (lit. “to sugar the strawberries”) ‘to suffer from nervous tremors’, constructed on the same pattern, the second can’t be passivized (15b).

- (15) a. 84 ans. À cet âge, tu l’imagines menu, éprouvant les pires difficultés à crapahuter dans les reliefs escarpés, bigleux et *sucrant les fraises*. (FrWac)
 “84 years old. At that age, you imagine him thin, with the worst difficulties to trudge in steep reliefs, squinting and *trembling*.”
- b. **Les fraises sont sucrées* par cet homme.
 (lit.) “Strawberries *are sugared* by this man.”

According to us, the blocking of this variation can be explained with the idiom’s meaning: SUCRER LES FRAISES denotes a perpetual behaviour,

¹² Verbs like AVOIR ‘to have’ and its quasi-synonyms like POSSÉDER, COMPORTER, COMPTER can’t be passivized; in the same way as COÛTER ‘to cost’ and its quasi-synonym VALOIR, or verbs like PESER ‘to weight’, etc. For a more complete list, see RIEGEL *et al.* (1994: 732).

caused by a sustainable state. We can also, here, apply the notion of *analyzability* discussed by Svensson (2008) and equivalent to Gibb's concept of *decomposability*:

Idioms like [...] *spill the beans*, and *lay down the law* are 'decomposable', because each component obviously contributes to the overall figurative interpretation. (Gibbs, 1994)

The example of SPILL THE BEANS 'to divulgate a secret' is also used by Nunberg *et al.* (1994) to show a correspondence between 'to spill' and 'divulgate' on the one hand, and between 'beans' and 'secret' on the other hand. This idea is shared by Osherson *et al.* (2009), who propose a semantic classification of idioms that takes in account analyzability.

In one sense, the whole meaning of NOYER LE POISSON can be scattered over its constituents, by analogy: NOYER 'to drown' is to 'to evade' what POISSON 'fish' is to 'issue'. This new meaning associated to POISSON allows a speaker to use it as the Rheme of a sentence like *le poisson a été noyé* 'the fish was drowned' (Pausé, in preparation)¹³. Even if SUCRER LES FRAISES is also based on a metaphor – it means 'to tremble (as if we were sugaring strawberries)' – we can't do any redistribution of the idiom's whole meaning on its constituents.

5. Conclusions

The phrasal status of idioms implies paradigmatic and syntagmatic variations which are not always equivalent to free phrases'. An accurate description of their lexico-syntactic characteristics is a precious asset, in particular for text generation. The choice of a dependency grammar allows to take into account the global overall syntactical organisation of the phrase, irrespectively of its linear organization. In other words, one becomes able to link a dependency structure to several possible linear organizations.

A resource based on the Explanatory Combinatorial Lexicology's framework offers a large database for research on idioms and their combinatorial properties. Once a linear LSS is associated to all idioms included in the base, we need to check for the paradigmatic and syntagmatic varia-

¹³ SEE MEL'ČUK (2001) about the opposition Theme/Rheme and its modelling in the Meaning-Text framework.

tions admitted. We have started this work with a sample of the most frequent idioms with simple syntactic patterns, like V Art NC. One interesting question is: how far the regular modifiers of a lexis can also modify this lexis as a constituent of an idiom – like *TAILLER UN COSTUME* which means ‘to criticize [...] a lot’ when *COSTUME* ‘suit’ is modified by *SUR MESURE* ‘tailor-maid’.

Acknowledgements

We are grateful to Raffaele Simone for his careful rereading of a first version of this paper, and to the anonymous reviewers for their valuable remarks.

Bibliography

- ABEILLÉ, A. (1995), *The flexibility of French idioms: A representation with Lexicalized Tree Adjoining Grammar*, in EVERAERT, M., VAN DER LINDEN, E.-J., ANDRÉ, S. and SCHREUDER, R. (1995, eds.), *Idioms: Structural and psychological perspectives*, Lawrence Erlbaum Associates, Hillsdale, pp. 15-42.
- BALLY, C. (1909), *Traité de Stylistique Française*. Vol. 1, Klincksieck, Paris.
- BARONI, M., BERNARDINI, S., FERRARESI, A. and ZANCHETTA, E. (2009), *The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora*, in «Language Resources and Evaluation», 43, 3, pp. 209-226.
- BURGER, H. (1998), *Phraseologie. Eine Einführung am Beispiel des Deutschen*, Erich Schmidt, Berlin.
- CROFT, W. and CRUSE, A. (2004), *From Idioms to Construction Grammar*, in CROFT, W. and CRUSE, A. (2004, eds.), *Cognitive Linguistics*, Cambridge University Press, Cambridge, pp. 225-256.
- FELLBAUM, C. (2014), *The syntax and grammar of idioms*, in KISS, T. and ALEXIA-DOU, A. (2014, eds.), *Handbook of Syntax*, De Gruyter, Berlin, pp. 776-802.
- FILLMORE, C. (2008), *Frame semantics meets construction grammar*, in BERNAL, E. and DECESARIS, J. (2008, eds.), *Proceedings of the XIII EURALEX International Congress*, Institut Universitari de Lingüística Aplicada, Barcelona, pp. 49-69.
- FILLMORE, C., KAY, P. and O'CONNOR, M. (1988), *Regularity and Idiomaticity in Grammatical Constructions: The case of «Let Alone»*, in «Language», 64, pp. 501-538.

- GADER, N., KOEHL, A. and POLGUÈRE, A. (2014), *A Lexical Network with a Morphological Model in It*, in ZOCK, M., RAPP, R. and HUANG, C.R. (2014, eds.), *Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon (CogALex)*, Association for Computational Linguistics, Stroudsburg, pp. 154-165.
- GERDES, K. and KAHANE, S. (2013), *Defining Dependencies (and Constituents)*, in GERDES, K., HAJIČOVÁ, E. and WANNER, L. (2013, eds.), *Computational Dependency Linguistics*, IOS Press, Amsterdam.
- GIBBS, R.W. (1994), *The Poetics of Mind: Figurative Thought, Language, and Understanding*, Cambridge University Press, New York.
- GLÄSER, R. (1998). *The Stylistic Potential of Phraseological Units in the Light of Genre Analysis*, in COWIE, A.P. (1998, ed.), *Phraseology: Theory, Analysis and Applications*, Oxford University Press, Oxford, pp. 125-143.
- GROSS, G. (1996), *Les expressions figées en français. Noms composés et autres locutions*, Ophrys, Paris.
- JORDANSKAJA, L. and MEL'ČUK, I. (2009), *Establishing an Inventory of Surface-Syntactic Relations: Valence-Controlled Surface-Syntactic Dependents of the Verb in French*, in POLGUÈRE, A. and MEL'ČUK, I. (2009, eds.), *Dependency in Linguistic Description*, John Benjamins Publishing Company, Amsterdam, pp. 151-234.
- KAHANE, S. (2012), *Why to Choose Dependency rather than Constituency for Syntax: A Formal Point of View*, in APRESJAN, J.D., BOGUSLAVSKY, I., L'HOMME, M.C., IOMDIN, L., MILIĆEVIĆ, J., POLGUÈRE, A. and WANNER, L. (2012, eds.), *Meanings, Texts, and Other Exciting Things: A Festschrift to Commemorate the 80th Anniversary of Professor Igor A. Mel'čuk*, Languages of Slavic Culture, Moscow, pp. 257-272.
- KAVKA, S. and ZYBERT, J. (2004), *Glimpses on the History of Idiomaticity Issues*, in «SKAZE Journal of Theoretical Linguistics», 1, pp. 54-66.
- KISS, G. R. (1968), *Words, Associations, and Networks*, in «Journal of Verbal Learning and Verbal Behaviour», 7, 4, pp. 707-713.
- LECLER, A. (2006), *Le défigement: un nouvel indicateur des marques du figement?*, in «Cahiers de praxématique», 46, pp. 43-60.
- LUX-POGODALLA, V. and POLGUÈRE, A. (2011), *Construction of a French Lexical Network: Methodological Issues*, in KOSEM, I. and KOSEM, K. (2011, eds.), *Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex 2011*, Institute for Applied Slovene Studies, Ljubljana, pp. 54-62.

- MEL'ČUK, I. (1998), *Collocations and Lexical Functions*, in COWIE, A.P. (1998, ed.), *Phraseology: Theory, Analysis and Applications*, Clarendon Press, Oxford, pp. 23-53.
- MEL'ČUK, I. (2012), *Phraseology in the Language, in the Dictionary, and in the Computer*, in COLSON, J.P. (2012, eds.), *Yearbook of Phraseology*, De Gruyter, Berlin, pp. 31-56.
- MEL'ČUK, I., CLAS, A. and POLGUÈRE, A. (1995), *Introduction à la lexicologie explicative et combinatoire*, Duculot, Paris / Louvain-la-Neuve.
- MEL'ČUK, I. and POLGUÈRE, A. (2016), *La définition lexicographique selon la lexicologie explicative et combinatoire*, in «Cahiers de lexicologie», 109, pp. 61-92.
- MILIĆEVIĆ, J. (2006), *A Short Guide to the Meaning-Text Linguistic, Theory*, in «Journal of Koralex», 8, pp. 187-233.
- MILLER, G.A., BECKWITH, R., FELLBAUM, C., GROSS, D. and MILLER, K.J. (1990), *Introduction to Wordnet: An On-Line Lexical Database*, in «International Journal of Lexicography», 3, 4, pp. 235-244.
- MOON, R. (1998), *Fixed Expressions and Idioms in English: A Corpus-Based Approach*, Oxford University Press, Oxford.
- NUNBERG, G., SAG, I.A. and WASOW, T. (1994), *Idioms*, in «Language», 70, 3, pp. 491-538.
- OSHERSON, A. and FELLBAUM, C. (2009), *The representation of idioms in WordNet*, in *Proceedings of ACL-IJCNLP 09*, Singapore.
- PAUSÉ, M.S. (in preparation), *Structure lexico-syntaxique des locutions du français et incidence sur leur combinatoire*, Thesis, Université de Lorraine, Nancy.
- PAUSÉ, M.S. and SIKORA, D. (2016), *Polysémie et homonymie des locutions dans un modèle lexical*, in NEVEU, F., BERGOUNIOUX, G., CÔTÉ, M.H., FOURNIER, J.M., HRIBA, L. and PRÉVOST, S. (2016, eds.), *Actes Du 5e Congrès de Linguistique Française*, Tours.
- POLGUÈRE, A. (2009), *Lexical Systems: Graph Models of Natural Language Lexicons*, in «Language Resources and Evaluation», 43, 1, pp. 41-55.
- POLGUÈRE, A. (2014), *From Writing Dictionaries to Weaving Lexical Networks*, in «International Journal of Lexicography», 27, 4, pp. 396-418.
- REY, A. and CHANTREAU, S. (2007), *Dictionnaire des expressions et locutions*, Le Robert, Paris.
- RIEGEL, M., PELLAT, J.C. and RIOUL, R. (2009, [1994]¹), *Grammaire méthodique du Français*, Presses Universitaires de France, Paris.

- RUPPENHOFER, J., ELLSWORTH, M., PETRUCK, M.R.L., JOHNSON, C.R. and SCHEFFCZYK, J. (2010), *FrameNet II: Extended Theory and Practice*, International Computer Science Institute, Berkeley.
- SVENSSON, M.H. (2008), *A Very Complex Criterion of Fixedness: Non-Compositionality*, in GRANGER, S. and PAQUOT, M. (2008, eds.), *Phraseology: An Interdisciplinary Perspective*, John Benjamins, Amsterdam / Philadelphia, pp. 81-93.
- YAKUBOVICH, Y. (2015), *Défigement dans les textes poétiques: typologie et exemples en français, espagnol, catalan, russe, bélarusse et polonais*, Thesis, Universitat Autònoma de Barcelona.

MARIE-SOPHIE PAUSÉ
Université de Lorraine and ATILF-CNRS
44 avenue de la Libération
54000 Nancy (France)
marie-sophie.pause@univ-lorraine.fr