

Is *Contamination* Good or Bad? A Corpus-assisted Case Study in Translating Evaluative Prosody

Mélanie Frank, Francesca Bartolesi, Silvia Bernardini, Alan Partington –
Dipartimento di Interpretazione e Traduzione, Università di Bologna, Campus di
Forlì

Citation: Frank, Mélanie, Francesca Bartolesi, Silvia Bernardini, Alan Partington (2020) “Is *Contamination* Good or Bad? A Corpus-assisted Case Study in Translating Evaluative Prosody”, in Adriano Ferraresi, Roberta Pederzoli, Sofia Cavalcanti, Randy Scansani (eds.) *Metodi e ambiti nella ricerca sulla traduzione, l’interpretazione e l’interculturalità – Research Methods and Themes in Translation, Interpreting and Intercultural Studies*, *MediAzioni* 29: B1-B28, <http://www.mediazioni.sitlec.unibo.it>, ISSN 1974-4382.

1. Evaluative prosody and cross-linguistic perspectives

As Sinclair (2004), among others, has shown, textual meaning does not depend solely on the sum of the meanings of individual words; instead, these meanings interact and alter once several words are brought together to form units of meaning. In Sinclair’s model, these units include a node word and its co-occurring words – referred to as “collocates” when they are statistically significant or simply very frequent. Thanks to the advent of large-scale corpora we have been able to observe that these “systematic regularities in the associations between sets of words, grammatical frames and particular meanings” (Biber 2012:10), are pervasive and form an integral part of our knowledge of a language.

These interactions between a node and its lexical environment can be very strong and salient, as is the case with idioms, proverbs and fixed expressions but, as pointed out by Stubbs (2001), other frequent combinations are more difficult to see, because they can vary in internal construction and level of fixedness – hence the notion of *semi-preconstructed* phrases (Sinclair 1991). These units create

nonetheless more or less conscious expectations in readers: when a word is encountered, other specific words are expected, because “[w]ords often make general predictions about the content of surrounding text”, and these predictions come from recurrent encounters (Stubbs 2001: 312). These expectations include those concerning the *evaluative* behaviour of lexical items, that is, whether they are normally used to express positive or negative evaluative meaning¹. We use “evaluation” in the bi-dimensional sense of “the indication of whether the speaker thinks that something (a person, thing, event, situation, idea etc.) is good or bad” (Thompson 1996: 65), with “good” and “bad” each existing of course in an infinite variety of ways.

Hoey (2005) named this psychological phenomenon *lexical priming*. This term describes the processes by which listeners, through repeated exposure, first internalise, and then reproduce, the constituent elements of language, their combinatorial possibilities and the semantic and pragmatic meanings associated with them. In natural language situations, lexico-grammatical items of different kinds (words, phrases, structures) are usually not met alone but in context. An item which co-occurs with another in discourse is thus recorded with it, and the co-occurrence becomes part of our knowledge of how both items tend to be employed within specific fields. The primings of the “same” item may well differ very radically, for instance, depending on whether it is found in, say, everyday conversation, in medical discourse or in political discourse, and so on. This knowledge is not limited to single co-occurring items but can extend to cases when an item co-occurs frequently with a series of items belonging to the same semantic field, for example, *fraught with* is generally followed by items indicating *danger, obstacles* and negative emotions. These, as it were, “preferred” semantic fields can be called the node item’s “semantic preferences” (Partington 2004). The term “preference” is a semi-humorous anthropomorphism, as if an item could actively “choose” its co-occurring preferences. The connotations of the co-occurring items and their semantic preferences contribute to building the

¹ The term used for expectations concerning evaluative meaning in Sinclair's model is *semantic prosody*. In this paper we use Partington's term *evaluative prosody* (2004: 131,150), which we believe is more transparent.

evaluative prosody of the word as primed in the speakers' minds. Put simply, we are primed to know whether items are normally used in positive or negative contexts of communication. We generally, again quite instinctively and unconsciously, combine arrays of items in texts which fit together evaluatively: a process Partington (2017) has termed "evaluative cohesion" or "harmony". In fact, our priming for evaluative use is so ingrained that we normally notice evaluative prosody only when our evaluative expectations are not met, either deliberately, for example, for dramatic or comic effects, or accidentally, due to production errors². Although we weave cohesive harmonies and employ the evaluative prosody of lexical items *instinctively* (Stewart 2010), as Xiao and McEnery (2006) point out, our *introspections* on patterns of collocation and evaluative prosody are limited and not always reliable, even in one of our native languages³. In other words, we are not always able to describe how items behave evaluatively, even in the language(s) we feel more confident with.

Moving on to a cross-linguistic perspective, numerous examples of prosodic mismatch between cross-language items considered by dictionaries as equivalents have been identified and analysed. In their cross-linguistic corpus-based study, Xiao and McEnery (2006) suggest that units of meaning and evaluative (semantic) prosody exist in all languages, however, they are unpredictable and do not always correspond across languages. Their identification is thus essential in the case of lexico-semantic variants and in order to distinguish among "near synonyms" in one language and seeming translation equivalents in another. For instance, Hunston (2007: 259) discusses her own perplexity at an email from a non-native English speaker thanking a supervisor for his "persistent help and advice"; the item *persistent* normally being primed for use in negative contexts. Partington (2018) notes how an extreme case across

² An example of the former could be "Mortal, guilty, but to me / The *entirely beautiful*" (Auden, *Lullaby*), given that corpus evidence suggests *entirely* is rarely followed by positive items (Partington 2017); an instance of the latter might be "technology has increased the number of positive-sum games (win-win relationships) that humans tend to be *embroiled in*" (Steven Pinker, TED Talks 2007, our emphasis).

³ We use the notion of "native speaker" for convenience, despite its controversial status (see e.g. Davies 2004).

Italian and English would be the translation of *egregio* (“outstanding” in the positive sense of “worthy”) for *egregious* (“outstandingly bad”).

Taking into account possible differences in semantic preference and evaluative prosody a translator/interpreter ideally needs to find the best-fitting equivalences among several alternatives in order to render the intended meaning correctly, depending on the context of situation. Yet language professionals, usually being non-native speakers of at least one of their working languages, are likely to have had less (or at least later) exposure to the priming processes of that language. Their access to, and awareness of, the evaluative behaviour of items may therefore be less than optimal (Wray 2002). Not only will their introspections be more limited, but their instinctive understanding or use of prosodies – and evaluative cohesion/harmony in general – may also be less assured, than in a native language (Aston 1999; Stewart 2005). This may well constitute in part the very definition of “(non-)nativeness”. This relative limited awareness could lead to a shift in the type of evaluative prosody employed, “[which] should be avoided in translation because it may create misunderstandings with respect to tone or content of the original message” (Berber Sardinha 2000:96). Oksefjell Ebeling (2013) uses bidirectional corpora to find out if a particular word in English, namely *cause*, and the most commonly used correspondences in Norwegian, share the same semantic prosody, showing that cross-linguistic equivalents found in dictionaries may have different prosodies across languages. Her study shows how corpus use can help translators to avoid errors induced by faulty assumptions about the generalizability of prosodies across languages, such as the use of the Norwegian word *forårsake* to translate *cause*, which may be considered a case of so-called “translationese” (Gellerstam 1986).

Concordances extracted from corpora may assist language professionals in their work, enabling them to observe repeated patterns and meanings in context and contributing to raising their awareness of the collocation and evaluative prosody of a word or unit of meaning (Xiao and McEnery, 2006). Partington (2004) argues that, as the prosody of an item can elude the individual's awareness and is not always indicated in dictionaries, it can sometimes be uncovered only upon analysing a considerable number of concordance lines, better still if these are

extended concordance lines which capture meanings at phrase or sentence levels.

In the remainder of this paper we illustrate these principles and methods using the word pair *contaminazione* (IT) / *contamination* (FR) as a case in point. In Section 2 we first examine the collocates, semantic preferences and evaluative prosody of *contaminazione* and *contamination* in two large general corpora of Italian and French, in order to check if the data found in lexicographic resources are confirmed by corpus evidence. We then present an exploratory corpus-based strategy to find translation equivalents in cases where semantic preferences and evaluative prosodies of a word and of its *prima facie* translation do not match. In Section 3 we conclude by summarizing our results concerning evaluative prosody and cohesion, which confirm the potential of multilingual corpus-assisted discourse methods for translation studies and translation practice.

2. Semantic contamination and evaluative prosody

2.1 Study aims

The aim of this case study is to establish, on the basis of corpus evidence, whether the French *contamination* overlaps in meaning and use with Italian *contaminazione* in all senses and contexts. Based on lexicographic evidence, there would be good reason to believe that this might be the case.

In the *De Agostini monolingual Italian dictionary* (1998)⁴ the entry for *contaminazione* has three senses⁵:

1. The diffusion of elements dangerous for human health or for life in a healthy environment, generally: *contamination of water and air*;
2. A mix of heterogeneous elements or elements of various origins in artistic or literary compositions: *contamination is typical of Latin comedies*;
3. in linguistics, the

⁴ *Il dizionario della lingua italiana*, Istituto Geografico De Agostini, Novara 1998.

⁵ In the interest of space, dictionary definitions are provided in English only (authors' translations).

intersection of two forms or structures which give origin to one new form or structure.

The *Zingarelli dictionary* (2016)⁶ adds a fourth meaning specifying that it also has figurative uses: “corruption, offence: contamination of other people’s innocence”.

For the French *contamination*, the monolingual French dictionary *Le Petit Robert* (2019)⁷ offers three definitions:

1. Blemish caused by impure contact
2. Invasion (of an object, environment, living organism) by pathogenic microorganisms [...] or by pollutants. The contamination of a person by an infectious agent. Water contamination by chemicals [...].
3. LING. [...] Contamination of a word by another or between two words.

The reference in the online dictionary of the *Académie Française*⁸ includes a fourth meaning related to the literary domain: “The alteration of a text because of interferences with one or more other texts. There has been contamination between these two excerpts of the Scripture. Contamination between these two manuscripts certainly happened in the Middle Ages”.

Interestingly, the TLFi (Trésor de la Langue Française informatisé)⁹, an online reference dictionary for teaching and research purposes, splits the various definitions into two sets: one with meanings conveying the idea of corruption or spreading of evil or damage, and one with meanings that do not convey such connotations and evaluations. This second category includes the definitions offered by the other dictionaries quoted above, as well as the following: “contamination of something by something or between something and something. Interpenetrations of two influences. *The contamination of Impressionism by Fauvism. The contamination between two recollections*” (Sartre, 1936: 107).

⁶ *Lo Zingarelli Vocabolario della lingua italiana*, Zanichelli editore, Bologna 2016.

⁷ *Le petit Robert de la langue française*, éditions Le Robert, Paris 2019.

⁸ Dictionnaire de l'Académie française informatisé, online: <https://academie.atilf.fr/>.

⁹ Trésor de la langue française informatisé, online: <http://atilf.atilf.fr/>.

We can thus summarise the different dictionary definitions of *both* words as follows: 1. the damaging of an environment, organism or other elements because of the introduction of a harmful substance, radioactivity or microorganisms; 2. the alteration of the characteristics of something (usually abstract), corruption, spreading of evil or vice; 3. the interpenetration of different influences (literature and art); 4. in linguistics, the influence of neighbouring elements on some specified element¹⁰.

Not surprisingly, most bilingual dictionaries only provide *contamination* as a translation equivalent for *contaminazione*, and vice versa. The *Garzanti dictionary* (1981)¹¹ explicitly states that the French translation for *contaminazione* is *contamination* in all its senses (“in tutti i significati”), and the *Le Robert & Signorelli dictionary* (2002)¹² specifies that the translation of *contaminazione* is *contamination* also in arts, literature and linguistics. The only exception we came across is the *Il Bosh dictionary* (2007)¹³, which also mentions “corruption” and “infection” alongside “contamination”. Based on extensive and authoritative dictionary evidence¹⁴, a translator would thus legitimately conclude that the two words are as perfect equivalents of each other as it gets. In the next sections, we intent to test this belief against corpus data.

¹⁰ Quite aptly, the subject of the present study itself – evaluative/semantic prosody – has often been defined as a form of contamination over time of the evaluation of co-occurring items on a linguistic item under study (Stewart 2010: 41-55).

¹¹ *Dizionario Garzanti francese-italiano/italiano-francese*, Garzanti editore, Milano 1981.

¹² *Le Robert & Signorelli Dictionnaire Français-Italien Italien-Français*, Dictionnaires Le Robert, Paris 2002.

¹³ *Il Bosh, Dizionario Francese-Italiano Italiano-Francese. Quinta edizione* Zanichelli editore, Bologna 2007

¹⁴ At the time of writing (25 August 2019), Google Translate and DeepL also give them as translation equivalents, DeepL offering “pollution” (FR) as an alternative.

2.2. Method

To (dis)confirm dictionary evidence, we studied the collocates, semantic preferences and evaluative prosody of *contaminazione* and *contamination* in two large pre-existing general corpora of Italian and French, the Araneum Italicum Maius, and the Araneum Francogallicum Maius (Benko 2014) The *Aranea* corpora were compiled by web crawling, with the aim of creating comparable corpora that can be used “for teaching purposes, but also in linguistic research (contrastive studies) and in lexicography (both mono- and bilingual)” (Benko, 2014: 248). The corpora were built following a standardised methodology: the web crawl was performed at (approximately) the same time, they are likely to contain similar text types, genres and registers, and are approximately of the same size (see Table 1). They seem therefore suitable for a comparative study.

	Araneum Italicum Maius	Araneum Francogallicum Maius
Language	Italian	French
Number of words	890,568,533	933,688,995
Number of texts	2,654,561	1,772,987
Date of crawling	2014	2013 and 2015

Tab. 1 Characteristics of the corpora used for the study.

We analysed the semantic preferences of the words in the two *Aranea* corpora, consulted through the SketchEngine application¹⁵. Following a lemma search for *contaminazione* and *contamination*, we obtained collocates for both nodes, sorting them by logDice score (the preferred collocational measure for large corpora according to the SketchEngine creators), within a span of five words to the left and to the right, with a frequency threshold of five occurrences in the whole corpus, and three occurrences in the specified span. We then examined the top 200 collocates for both nodes and categorised them according to their

¹⁵ www.sketchengine.co.uk.

semantic field, ranking the results by frequency and matching them cross-linguistically.

Second, we analysed a sample of 200 concordance lines for *contaminazione* and *contamination* randomly extracted from each corpus, to see if the richer co-textual analysis afforded by the concordance tool confirmed the tendencies highlighted by the collocation tool regarding the evaluative prosody of the two words. Six concordance lines were discarded from the Italian sample and two from the French one. When a corpus is crawled from the web, automatic procedures are used for language identification, de-duplication, cleaning, etc. On occasion, therefore, one sometimes finds that a small number of concordance lines are not in the right language, contain machine-generated text, and so on. This was the case with eight concordance lines out of the 400 we randomly sampled from the two corpora. The collocates and concordances were analysed by one rater and, whenever needed for the analysis (for example, when two or more interpretations were possible), the wider context was checked to ascertain the meaning and use of the words or lines under study. The prosodies (negative, neutral, positive) were evaluated taking into account the meaning of the words surrounding the node, and the general meaning conveyed by the co-text, following the procedure described in Sinclair (2004).

2.3. Results

2.3.1. Contamination as environmental and microbiological hazards

The Italian corpus contains 10,567 occurrences (8.81 per million words, henceforth pmw) of the lemma *contaminazione*. The French corpus contains 10,636 occurrences (8.86 pmw) of the lemma *contamination*. Their relative frequencies in the two corpora are thus very similar. Figures 1 and 2 show the first 50 collocates of the lemmas *contaminazione* and *contamination* in each corpus. Unsurprisingly, in both lists many have a negative connotation (e.g. *radioattivo/radioactive*, *rischio/risqué*, *fecale/fécal*) and are linked to the semantic field of environmental hazards and microbiology. These results illustrate the first

definition of the two words provided by all the dictionaries we consulted (see 2.1), that is, damage and alteration of something's characteristics.

	Cooccurrence count	Candidate count	T-score	MI	logDice
P N genere genero	224	16,178	14.957	10.618	8.100
P N radioattivo	109	4,340	10.436	11.477	7.904
P N batterico	111	5,049	10.531	11.285	7.863
P N crociato	67	3,828	8.181	10.956	7.252
P N falda	80	7,193	8.937	10.302	7.205
P N microbico	50	1,157	7.069	12.260	7.126
P N microbiologico	49	1,285	6.998	12.080	7.081
P N accidentale	57	4,049	7.545	10.642	6.997
P N glutine	75	8,856	8.651	9.909	6.983
P N acquifero	43	2,078	6.554	11.198	6.799
P N diossina	40	2,070	6.321	11.099	6.696
P N reciproco	99	21,433	9.930	9.034	6.663
P N ibridazione	33	1,063	5.742	11.783	6.538
P N fecale	32	727	5.655	12.287	6.536
P N micotossine	29	402	5.384	13.000	6.436
P N OGM	40	4,664	6.318	9.927	6.427
P N rischio	504	182,370	22.378	8.293	6.419
P N stile stilo	81	21,634	8.978	8.731	6.365
P N contaminazione	53	10,567	7.267	9.153	6.360
P N linguaggio	205	72,964	14.272	8.317	6.329
P N suolo	102	31,123	10.072	8.539	6.325
P N alimento	134	45,474	11.541	8.386	6.291
P N arsenico	28	1,700	5.288	10.869	6.224
P N jazz	94	31,872	9.666	8.387	6.181
P N inquinamento	71	22,511	8.402	8.484	6.136
P N batterio	45	11,379	6.693	8.810	6.070
P N saperi	34	6,720	5.820	9.166	6.010
P N sperimentazione	86	33,580	9.241	8.184	5.996
P N chimico	94	38,799	9.660	8.103	5.963
P N arte arta	96	41,194	9.760	8.047	5.925
P N evitare	300	156,246	17.241	7.768	5.880
P N esente	31	7,088	5.556	8.956	5.846
P N microrganismo	25	3,710	4.993	9.579	5.842
P N stilistico	34	9,109	5.817	8.727	5.823
P N sonorità	30	7,161	5.465	8.894	5.793
P N incrociare	48	18,612	6.904	8.194	5.752
P N uranio	22	3,179	4.684	9.618	5.712
P N intreccio	32	9,752	5.641	8.541	5.689
P N blues	35	11,668	5.898	8.412	5.688
P N potabile	26	6,444	5.087	8.839	5.646
P N prevenire	60	28,882	7.713	7.882	5.639
P N ambientale	195	120,907	13.888	7.516	5.602
P N sotterraneo	35	13,588	5.895	8.192	5.569
P N pesticida	21	4,048	4.574	9.202	5.557
P N etnico	33	12,453	5.725	8.233	5.553
P N aflatossine	15	165	3.872	13.333	5.517
P N mescolanza	17	1,660	4.119	10.183	5.509
P N artistico	160	106,468	12.574	7.414	5.485
P N impoverire	20	4,237	4.463	9.066	5.468
P N pop	42	20,566	6.452	7.857	5.466

Fig.1 The first 50 collocates of *contaminazione* in the Italian corpus, sorted by logDice.

	Cooccurrence count	Candidate count	T-score	MI	logDice
P N radioactif	234	7,488	15.292	11.783	8.724
P N virus	292	30,234	17.072	10.089	7.871
P N bactérien	99	3,595	9.946	11.601	7.832
P N VIH	177	15,774	13.293	10.306	7.778
P N microbien	77	1,582	8.773	12.422	7.690
P N risque	1,300	226,456	35.999	9.339	7.489
P N microbiologique	62	970	7.872	12.816	7.451
P N contamination	108	10,636	10.383	10.161	7.378
P N fécal	52	953	7.209	12.587	7.199
P N croisé	70	6,208	8.360	10.313	7.089
P N accidentel	63	4,559	7.932	10.606	7.085
P N nappe	79	9,850	8.878	9.821	6.981
P N phréatique	48	2,117	6.925	11.320	6.946
P N OGM	79	11,124	8.877	9.646	6.894
P N prévenir	216	51,878	14.665	8.875	6.823
P N pesticide	79	13,350	8.874	9.382	6.753
P N chimique	149	35,353	12.180	8.893	6.730
P N éviter	559	169,279	23.579	8.541	6.669
P N bactérie	77	14,900	8.759	9.187	6.626
P N souterrain	78	15,770	8.815	9.124	6.596
P N bactériologique	33	954	5.743	11.930	6.543
P N viral	50	7,241	7.061	9.605	6.518
P N sédiment	39	3,463	6.240	10.311	6.502
P N aliment	171	52,560	13.041	8.519	6.470
P N croiser	163	51,228	12.731	8.487	6.431
P N éventuel	156	48,749	12.455	8.496	6.427
P N coli	30	1,093	5.475	11.596	6.389
P N sida	72	17,944	8.466	8.822	6.367
P N résiduel	37	5,298	6.075	9.621	6.249
P N mercure	34	4,190	5.824	9.838	6.231
P N VHC	27	1,459	5.193	11.027	6.192
P N pathogène	33	4,437	5.737	9.712	6.164
P N potable	47	12,008	6.840	8.786	6.087
P N ingestion	27	2,609	5.191	10.189	6.061
P N inhalation	25	2,103	4.996	10.389	6.006
P N contaminer	44	11,881	6.617	8.706	6.000
P N génétique	69	24,922	8.280	8.287	5.990
P N hépatite	33	6,414	5.734	9.181	5.986
P N dioxine	23	1,282	4.793	10.983	5.982
P N propagation	33	6,714	5.734	9.115	5.961
P N toxicité	28	4,086	5.284	9.594	5.961
P N source	360	188,419	18.885	7.751	5.889
P N césium	20	628	4.470	11.811	5.862
P N minimiser	34	9,442	5.816	8.666	5.794
P N PCB	20	1,217	4.469	10.856	5.788
P N Fukushima	26	5,016	5.090	9.191	5.766
P N E.	38	12,306	6.146	8.444	5.762
P N germe	27	6,085	5.185	8.967	5.725
P N détecter	53	22,246	7.253	8.070	5.722
P N microorganisme	19	1,217	4.356	10.782	5.714

Fig. 2 The first 50 collocates of *contamination* in the French corpus, sorted by logDice.

Extending the analysis to the top 200 collocates, words that belong to this semantic field represent 55.5% of the 200-top list of the collocates of *contaminazione* and 95.5% of the 200-top list of collocates of *contamination*. They can in turn be categorised into four subsets (see Tables 2 and 3). The first set contains words referring to agents responsible of contamination (such as *pesticides* or *bacteria*), the second set includes words referring to elements that can be contaminated (such as the *environment*, or *food products*), the third set contains adjectives describing the type of contamination (such as *microbiological* or *accidental*) and the fourth set contains words that refer to the presence of a danger (such as *pollution*) or the need to protect something from a danger (such as the verb *preserve*). These four subsets account for 94% of the collocates of *contamination* and for 92% of the collocates of *contaminazione* that are part of the semantic field of environmental hazards and microbiology. The collocates that

refer to notions and objects found both in Italian and French are highlighted in bold in Tables 2 and 3.

Sources (32%)	Objects (15%)	Types (21%)	Danger (24%)
<i>glutine</i> (gluten)	acqua (water)	ambientale (environmental)	rischio (risk)
OGM/Ogm/ogm (GMO)	alimento (food)	batterico (bacterial)	evitare (to avoid)
sostanza (substance)	suolo (soil)	crociato (cross)	inquinamento (pollution)
metallo (metal)	falda (ground water)	radioattivo (radioactive)	<i>pericolo</i> (danger)
diossina (dioxin)	<i>terreno</i> (land)	alimentare (food)	prevenire (to avert)
<i>microrganismo</i> (micro-organism)	cibo (food product)	chimico (chemical)	<i>preservare</i> (to preserve)
batterio (bacteria)	<i>sorgente</i> (spring)	accidentale (accidental)	impedire (to prevent)
pesticida (pesticide)	coltivazione (growing)	microbiologico (microbiological)	<i>tossico</i> (toxic)

Tab. 2 Subsets of Italian collocates of *contaminazione* that are part of the semantic field of environmental hazards and microbiology in the top 200 collocates, with examples. The percentages refer to collocate types.

Sources (28%)	Objects (13%)	Types (25%)	Danger (28%)
<i>virus</i> (<i>virus</i>)	eau (<i>water</i>)	croisé (<i>cross</i>)	risque (<i>risk</i>)
<i>VIH</i> (<i>HIV</i>)	aliment (<i>food</i>)	radioactif (<i>radioactif</i>)	éviter (<i>to avoid</i>)
OGM/Ogm/ogm (<i>GMO</i>)	<i>environnement</i> (<i>environment</i>)	chimique (<i>chemical</i>)	prévenir (<i>to avert</i>)
pesticide (<i>pesticide</i>)	sol (<i>soils</i>)	alimentaire (<i>food</i>)	<i>réduire</i> (<i>to reduce</i>)
bactérie (<i>bacteria</i>)	nappe (<i>ground water</i>)	bactérien <i>bacterial</i>	<i>limiter</i> (<i>to limit</i>)
métal (<i>metal</i>)	lait (<i>milk</i>)	microbiologique (<i>microbiological</i>)	pollution (<i>pollution</i>)
substance (<i>substance</i>)	<i>viande</i> (<i>meat</i>)	accidentel (<i>accidental</i>)	empêcher (<i>to prevent</i>)
dioxine (<i>dioxin</i>)	denrée (<i>food products</i>)	environnemental (<i>environmental</i>)	<i>protéger</i> (<i>to protect</i>)

Tab. 3 Subsets of French collocates of *contamination* that are part of the semantic field of environmental hazards and microbiology in the top 200 collocates, with examples. The percentages refer to collocate types.

In Tables 2 and 3, one may see that some collocates seem to be more specifically linked to one of the two languages, like *virus* and *VIH* (appearing among the top collocates of *contamination*, with no corresponding Italian terms among the first 200 collocates of *contaminazione*), and *glutine* (the ninth collocate of *contaminazione*, with no corresponding French term among the first 200 collocates of *contamination*). Despite these particularities, the percentages each subset accounts for, as far as the semantic field of environmental hazards and microbiology is concerned, are quite similar both in French and in Italian, and

many collocates are present in both lists, suggesting that *contaminazione* and *contamination* behave quite similarly when used with in their literal sense.

2.3.2. Figurative meanings: *contaminazione* and *contamination* as cultural influence and as interpenetration?

A look at the list of Italian collocates in Figure 1 shows that a set of words stand out which distinguish Italian *contaminazione* from French *contamination*. These are part of the semantic field of culture, art and music: *jazz, stilistico, blues, etnico, artistico, pop* and *sonorità*. Once again, extending the analysis to the top 200 collocates, other words pop up that belong to this semantic field: *generi, culturale, linguaggio, cultura, musica, artistico, musicale, arte, arti, tradizione, stili, contemporaneo, rock, genere, stile, visivo, danza, espressivo, sonoro, saperi, creatività, linguistico, pittura, folk, tango, funk*. Altogether, there are 31 words in this set, i.e., 15.5% of the top 200 collocates of *contaminazione*.

The examination of the first 200 collocates in the Italian list highlights two other interesting facts. The first is the presence of 12 words (6%) referring to the third meaning provided by the dictionaries (“interpenetration”, see Section 2.1), namely: *reciproco, scambio, influenza, ibridazione, intreccio, condivisione, interazione, ispirazioni, incrocio, mescolanza, suggestione, meticciano*. The second is the presence of 7 adjectives (3.5%) conveying positive evaluation: *creativo, positivo, virtuoso, fecondo, fertile, proficuo, fruttuoso*. A follow-up concordance analysis shows that most of these adjectives are directly connected with the noun *contaminazione*, either pre-modifying or post-modifying it (108 out of 154 occurrences). Furthermore, most of the concordance lines (147 out of 154) are instances of the figurative use, with a semantic preference for the field of culture/arts/music, as discussed above in this subsection. In the few cases in which (seemingly) positive adjectives (such as *positive* and *fertile*) co-occur with *contaminazione* used in its literal meaning, they do not contribute to a positive evaluative prosody. In example 1, *positive* refers to medical tests (“tested positive to faecal contamination”), while in example 2 *fertile* refers to soil that is “free from” contamination / contaminants.

1. di E. coli come segno di contaminazione fecale. Sono risultati **positivi** per **contaminazione** fecale prodotti provenienti da ogni città e da ciascuna catena di negozi.

[of E. coli as a sign of faecal contamination. Products coming from every city and from every chain store tested positive for faecal contamination.]

2. dell'agricoltore assicurare un terreno **fertile** e tuttavia privo di **contaminazioni**, per far sì che le mele abbiamo poi la classificazione di biologico.

[of the farmer to guarantee a fertile soil free from contaminants, so that apples can be then classified as biological.]

Moving on to the analysis of the French data, the only item in the collocate list which could refer to the concept of interpenetration is the adjective/past participle *croisé*. This is however always used in the technical expression “*contamination croisée*” (cross-contamination), with the meaning of “Transfer of a contaminant from a source, specimen, etc., to a different or uncontaminated one” (Oxford English Dictionary, [OED])¹⁶.

To summarise, not a single word referring to culture/arts/music was found in the French collocate list, nor any word with a clear positive connotation, suggesting that the semantic preferences and prosody of *contaminazione* and *contamination* may overlap only partially. We will return to the implications of these observations in Section 3 below.

2.3.3. Concordance analysis

The collocate tool offers a picture of “the words which occur in the neighbourhood of your search word” (Scott, Wordsmith V8) and, since the collocational display includes both content and grammatical lexis, can often also offer an overview of the lexical profile of an item, including clues to its syntagmatic behaviour. Yet to really get a feeling for how words are used in connected text, and to confirm

¹⁶ Oxford English Dictionary (OED) 1989-2020. Consulted online: <https://www.oed.com>.

results obtained from collocation analysis, browsing concordances is essential. Out of the 194 randomly selected concordance lines from the Italian corpus, 73 deal with environmental and microbiological hazards (37.6%; example 3). The evaluation conveyed is neutral/positive in 1 case out of 73 (example 4), and negative in the remaining 72.

3. estrattiva interferisce con il deflusso delle acque di falda e determina **contaminazioni** pericolose per lo stato di qualità delle acque superficiali e sotterranee.

[mining industry interferes with the flow of ground water and it results in contaminations dangerous for the quality of surface and ground water.]

4. supportare i processi naturali di biodegradazione dei composti organici di **contaminazione** ad opera dei microrganismi presenti nel sottosuolo

[support the natural processes of organic compounds biodegradation done by the micro-organisms present in the subsoil]

In 121 lines (62.3%), *contaminazione* is used with a figurative meaning. The evaluation is neutral in 14 cases, negative in 22, and positive in 85. The semantic preference for culture/arts/music is observed in 96 concordance lines (example 5); example 6 illustrates a partially distinct use, from philosophy.

5. cinquanta opere, la città, i colori, i santi, la storia rivelano tutte le **contaminazioni** artistiche di un porto mediterraneo aperto all'Europa e all'oriente

[fifty masterpieces, the city, its colours, its saints, its history reveal all the artistic influences of a Mediterranean port open to Europe and to the East]

6. penetra nel visibile e al contempo si lascia penetrare da esso in una logica di **contaminazione** tra il vedente e il visibile c'è un chiasmo, una relazione, simile a quella

[penetrates into the visible and lets itself be penetrated by it at the same time in a logic of mutual influence between those who see and what is seen there is a chiasm, a relationship, similar to that]

Zooming in on the culture/arts/music semantic field, the evaluative prosody is overwhelmingly positive, with 78 concordance lines conveying positive evaluation (7 are negative and 11 neutral). Positivity is conveyed by adjectives (e.g. *artistico*,

aperto, creativo, nuovo, riuscito, meraviglioso), nouns (e.g. *passione, evoluzione, viaggio, scambio, dialogo, collaborazione*) and verbs (*stimolare, creare, arricchire, imparare, apprezzare, nutrire*). Lexical items in the co-text provided evidence of a positive evaluative prosody of the overall meaning when the words for which the positivity may be less unequivocal (as *artistico, viaggio* or *dialogo*) were employed. In the 25 concordance lines in which *contaminazione* is used figuratively but does not refer to the semantic field of culture/arts/music, the prosody is more often negative (16 cases, example 7), with 7 positive (example 8), and 2 neutral cases (example 9).

7. Così come non lo sono le collusioni tra politica ed industria e le **contaminazioni** tra malavita ed alta finanza, che fanno solitamente da cornice a squallide storie condite [...]

[Nor are the collusions between politics and industry and the contamination between organised crime and high finance, which usually come with sleazy stories spiced with. ...]

8. L'impegno con il mondo della scuola In un'ottica di **contaminazione** e collaborazione continua e con il coinvolgimento degli enti e delle istituzioni del [...]

[The commitment with the world of education with a view to cooperation and collaboration continues and with the involvement of the bodies and institutions of the ...]

9. nuova categoria con la quale classificare i regimi in transizione ovvero si tratta di una **contaminazione** – una “zona grigia” – tra generi dicotomici? Il concetto di regime ibrido

[new category which permits us to classify the transitional regimes, that is, a combination – a "grey area" – of dichotomous genres? The concept of hybrid regime]

The sample of 198 concordance lines from the French corpus contains 191 lines linked to the semantic field of environmental and microbiological hazards (96.4%); the evaluative prosody is consistently negative (see example 10).

10. appliquée à des problèmes tels que le changement climatique, la gestion des déchets et la **contamination** des eaux souterraines. Biographie: Dr. Jack Cornett est professeur à [...]

[applied to issues such as climate change, waste management and groundwater contamination. Biography: Dr. Jack Cornett is a professor at ...]

In the 7 remaining concordance lines in which *contamination* is used figuratively, the evaluation conveyed is also consistently negative (example 11). In only one case does the word refer to the semantic field of culture (example 12). In this case, the evaluation appears somewhat difficult to judge, yet the word *contamination* is part of a list of nouns that describe a literary work, culminating with the clearly negative word *trahison* (“betrayal”), emphasized by the exclamation point.

11. demandant de me purifier. Je sais que Toi et Toi seul es capable de renverser le processus de **contamination**. Rends-moi pur, pardonne-moi mes péchés afin que je puisse être une nouvelle

[asking to purify me. I know that You and You alone are capable of reversing the process of contamination. Make me pure, forgive my sins so that I can be a new]

12. il le dit lui-même : « Traduction, translation, actualisation, imitation, **contamination**, réécriture, re-visitation, réinvention ou bien même trahison ! au lecteur

[as he says himself: “translation, transposition, updating, imitation, contamination, rewriting, retelling, reinvention or even betrayal! it is up to the reader]

Figures 3 and 4 summarise the results of the concordance analysis, contrasting evaluative prosodies and semantic preferences across the Italian and the French data.

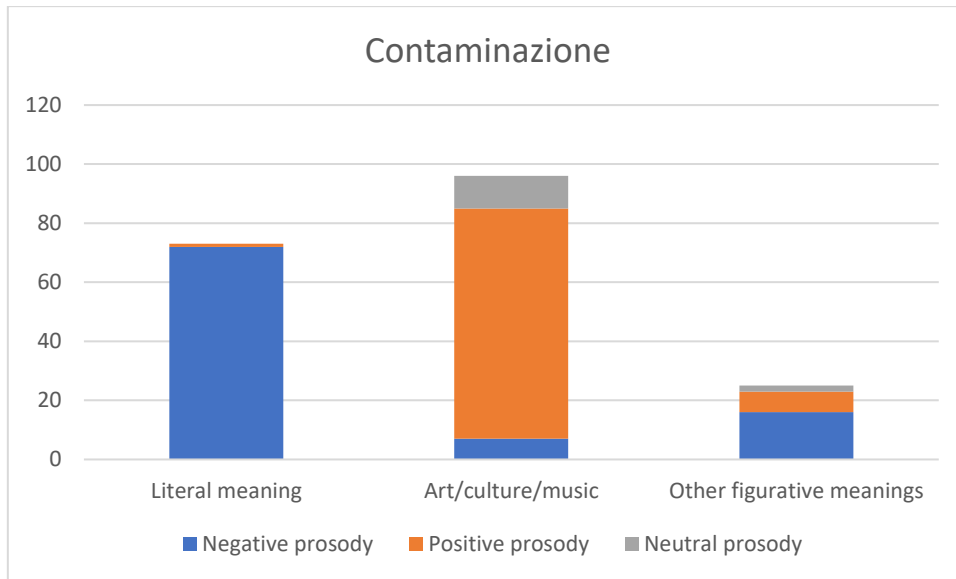


Fig. 3 Evaluative prosody and semantic preferences of *contaminazione* (concordance analysis).

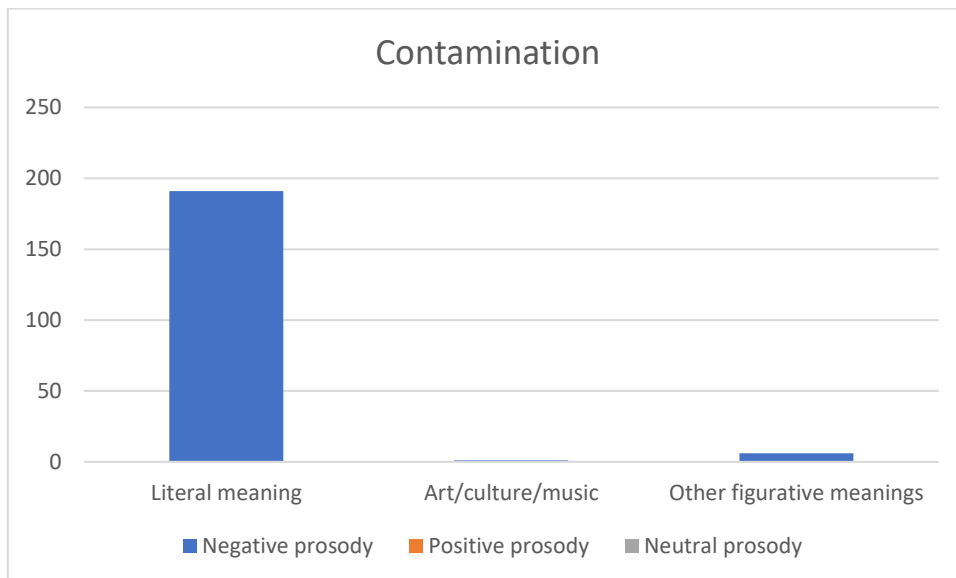


Fig. 4 Evaluative prosody and semantic preference of *contamination* (concordance analysis).

The more detailed analysis afforded by the concordance tool has confirmed the insights gleaned from the collocation analysis. The results regarding *contaminazione* and *contamination* used in their literal meaning are very similar in both languages with respect to the contexts of use, the way they collocate with words that are part of the semantic fields of environmental hazards and

microbiology and their evaluative prosody. When used with a figurative meaning, however, things change. Almost half of the collocates and more than half of the sample of concordance lines we analysed for *contaminazione* show a figurative use; in contrast, with *contamination*, the percentage falls below 5%. More importantly, whilst the evaluative prosody in the literal use seems overwhelmingly negative in both languages, in the figurative use *contaminazione* has mostly a positive evaluative prosody. In French, no positive prosody was ever observed in the data. The positive evaluative prosody of *contaminazione* seems to be especially linked to the semantic field of culture/arts/music, which is completely absent in the results related to *contamination*.

The obvious question a translator would ask at this point is: since *contamination* is not normally used as an equivalent for *contaminazione* in the positively evaluating sense related to music/arts/culture, and since dictionaries do not offer much in the way of alternative solutions, can corpus methods help find a more appropriate equivalent? Before moving on to our final section, we briefly illustrate an exploratory attempt at using corpus evidence to answer this question. Due to space limits, we will only sketch a possible method here, to be explored further in future work.

2.4. Follow up: an exploratory collocate-driven strategy for a cross-linguistic equivalence search

On the basis of extensive collocation and concordance analysis, we came to the conclusion that *contaminazione* (IT) and *contamination* (FR) share the same semantic preference and the same negative evaluative prosody with respect to the semantic field of environmental and microbiological hazards. Instead, the semantic preference shown by *contaminazione* for the fields of culture, arts and music, especially with regard to the mingling of one influence with others, very often with a positive evaluative prosody, is not associated with *contamination* in our corpus. In a follow-up exploratory study, we therefore attempted to develop a strategy that would allow us to retrieve one or more French equivalents of *contaminazione* in the sense(s) for which *contamination* would not seem to be

adequate. For this purpose, we tested a collocate-driven procedure consisting of three main steps.

First, we selected and translated a number of Italian collocates of *contaminazione* into French, to be used as nodes (or “seeds”) in new searches of the French corpus. These were manually chosen from among: 1. the set of collocates of *contaminazione* belonging to the semantic fields of culture/arts/music; 2. the set of collocates of *contaminazione* belonging to the semantic field of interpenetration; 3. the positive adjectives collocating with *contaminazione*. Table 4 lists the chosen Italian collocates and their translations. This procedure was largely manual and intuitive: we tried to identify typical yet sufficiently general words, avoiding, for instance, specific music genres. In future work we would like to try and experiment with methods for choosing the seed collocates in a corpus-driven fashion, for instance, relying on statistical association measures. In addition, translation was performed manually, and focused on individual collocate words. Since we assume that meaning is spread over units larger than the single word, this procedure is clearly less than ideal. In this exploratory phase, however, we use these translations as rough-and-ready approximations.

Set	Italian collocates	French translation (new node)
Culture/arts/music	<i>generi, musica, linguaggio, cultura, arte, stili, culturale, musicale, artistico, saperi</i>	<i>genres, musique, langage, culture, art, styles, culturel, musical, artistique, savoirs</i>
Interpenetration	<i>reciproco, scambio, influenza, condivisione, interazione, mescolanza, suggestione, incrocio, ispirazione</i>	<i>réciproque, échange, influence, partage, interaction, mélange, suggestion, croisement, inspiration</i>
Positive adjectives	<i>creativo, positive, virtuoso, fecondo, fertile, fruttuoso</i>	<i>créatif, positif, vertueux, fécond, fertile, fructueux</i>

Tab. 4 French dictionary equivalents of collocates of *contaminazione*.

Second, we carried out lemma searches for each one of the new nodes in the *Araneum Francogallicum Maius* corpus. We then collected collocates, sorted by

logDice score, within a span of two words to the left and two to the right, with a frequency threshold of 10 occurrences in the whole corpus, and 7 occurrences in the specified span. The thresholds are higher than in the first collocate analysis performed in this study (Section 2.2), so as to favour precision over recall. Given that this research phase is exploratory, it was deemed important that the amount of retrieved evidence would not be overwhelming.

Third, we copied the collocates into three Excel sheets, corresponding to the three seed sets, one column per seed, and counted the collocates common to two or more columns. Our hypothesis was that candidate equivalents for *contaminazione* could be found by browsing the collocates shared by several seeds. To do this, we counted the number of times that a given cell content is repeated within a given range¹⁷. We created matching columns that show the results of this count, namely, how many times a given word appears in a given seed set. Table 5 shows the top six collocates from the *positive adjective* sheet. For reasons of space it is not possible to provide the complete view of the sheet. Focusing on the last column, we can see that the adjective *fructueux* (our proposed French translation for Italian *fruttuoso*, in turn one of the positive adjectives collocating with *contaminazione*) collocates with *collaboration*, *échange*, *avérer*, *coopération*, *enrichissant* and *dialogue*. When the number in parentheses is higher than one, the collocate in the corresponding cell also collocates with one or more of the other seeds (*créatif*, *positif*, *vertueux*, *fécond* or *fertile*). This is the case with *dialogue* and *échange*, both found four times in the lists.

¹⁷ We used the following MS Excel formula (with Excel 2016 set up in English): =COUNTIF(\$A\$2:\$I\$1001;A2), where A2 corresponds to the first cell of the first column and I1001 corresponds to the last cell of the last column, variable across sheets.

Seed	créatif	positif	vertueux	Fécond	fertile	Fructueux
coll.	<i>loisir</i> (1)	<i>néгатif</i> (1)	<i>cercle</i> (1)	<i>multipliez-vous</i> (1)	<i>terreau</i> (2)	<i>collaboration</i> (3)
coll.	<i>atelier</i> (1)	<i>impact</i> (1)	<i>indignation</i> (1)	<i>extraordinairement</i> (1)	<i>imagination</i> (3)	<i>échange</i> (4)
coll.	<i>ludique</i> (1)	<i>bilan</i> (1)	<i>spirallspirale</i> (2)	<i>Imagination</i> (3)	<i>plaine</i> (1)	<i>avérer</i> (4)
coll.	<i>processus</i> (2)	<i>discrimination</i> (1)	<i>écologiquement</i> (1)	<i>terreau</i> (2)	<i>croissant</i> (1)	<i>coopération</i> (2)
coll.	<i>inventif</i> (1)	<i>dénivelé</i> (1)	<i>obéissant</i> (1)	<i>dialogue</i> (4)	<i>limon</i> (1)	<i>enrichissant</i> (1)
coll.	<i>potentiel</i> (1)	<i>attitude</i> (2)	<i>vicieux</i> (1)	<i>extrêmement</i> (5)	<i>rebondissement</i> (1)	<i>dialogue</i> (4)

Tab. 5 The first six collocates from the *positive adjective* Excel sheet, with number of occurrences of each collocate across the various columns.

On the basis of this procedure, which was repeated for the three sheets, we obtained the list of French words related to the notion of Italian *contaminazione*, used in a figurative sense. These are shown in Table 6 below.

French words related to Italian "contaminazione"	Culture / Arts / Music category (times)	Interpenetration category (times)	Positive adjectives category (times)
<i>mélange (n.)</i>	3	3	0
<i>échange (n.)</i>	3	4	4
<i>influence (n.)</i>	3	6	1
<i>partage (n.)</i>	0	3	0
<i>dialogue (n.)</i>	3	1	4
<i>rencontre (n.)</i>	3	2	4
<i>interaction (n.)</i>	0	3	3
<i>inspiration (n.)</i>	2	4	2
<i>lien (n.)</i>	5	3	2
<i>collaboration (n.)</i>	2	3	3
<i>carrefour (n.)</i>	1	3	0
<i>relation (n.)</i>	1	4	3
<i>mélanger (v.)</i>	2	3	0
<i>mêler (v.)</i>	4	2	0
<i>partager (v.)</i>	1	5	1
<i>réci-proque (adj.)</i>	0	6	0
<i>multiple (adj.)</i>	5	6	2
<i>mutuel (adj.)</i>	0	6	0
<i>divers (adj.)</i>	6	7	2
<i>entre (prep.)</i>	6	7	5

Tab. 6 Candidate French words related to Italian *contaminazione* (figurative) subdivided into categories.

As Table 6 reveals, several collocates appearing in one semantic category are also shared by others. For example, *échange* appears more than once in all the categories. We believe that this aspect might be seen as a further indication that these collocates could be (or point to) good candidate equivalents.

We would suggest that a translator from Italian into French could draw considerable insight from a procedure such as this one in her search for a context-sensitive equivalent of *contaminazione*, both relying on her (now augmented) intuition or using it as a starting point for further corpus searches. In this work, we have concentrated on the second of these processes. Since the presence of the optional determiner *entre* ("between") might suggest the syntactic pattern NOUN+ADJ+ENTRE+NOUN, in which the first noun could be a potential translation for *contaminazione*, we carried out a search for an adjective (from among the set above: *créatif/positif/vertueux/fécond/fertile/fructueux*), followed by *entre* and a noun (*genre/musique/style/culture/savoir/langage/art*). This procedure returned more potential equivalents in context, such as *alliance*, *collision*, *symbiose*, *va et vient*.

3. Discussion and conclusions

The present corpus-assisted study has shown that the Italian word *contaminazione* and its French cognate *contamination* share the same semantic preference and the same negative evaluative prosody as far as the semantic field of environmental and microbiological hazards is concerned. But it seems that another context of use of *contaminazione*, that of culture (including literature, arts and music), is not shared by the French word *contamination* in our corpus. In the context of culture, the item *contaminazione* displays a very positive evaluative prosody. The dictionaries we consulted give no information on these two distinct evaluative prosodies and can induce the reader to think that the two words are perfect equivalents in all contexts. It is easy to understand how a translator can be misled in a case like this where 1) the two words are cognates, 2) they have corresponding definitions, 3) the dictionaries seem to imply that they have the same semantic preferences and evaluative prosodies in all contexts.

An explanation for the reason why no trace of the specific use linked to interpenetration in literature/arts/artistic domains was found in the French corpus, despite it being mentioned in monolingual French dictionaries, could be that *contamination* in this sense has remained a "technical" term used only by a restricted circle of experts, whereas the Italian equivalent may have evolved to become part of the common vocabulary used to enhance positive influences in the cultural sector. The sample of French occurrences with a figurative meaning is too small to draw conclusions about its specific evaluative prosody, but we would argue that, since *contamination* in 95% of cases expresses the highly negative evaluation associated with environmental or microbiological hazards, the French reader will be overwhelmingly influenced by this negative connotation even when it is used in a figurative meaning. Hence the importance of not translating *contaminazione* as *contamination* in this specific context.

While we only studied one word, we believe that the mismatch observed in evaluative prosodies of figurative and non-figurative uses may be a more general phenomenon. Partington (2017) shows how differently the English item *orchestrat** behaves when employed literally or figuratively. When used literally as “to combine harmoniously the instruments in an orchestra” (OED), it co-occurs with very positive items such as *astute and attractive, wonderful discoveries, luxurious*, but when used figuratively in the field of politics, its evaluative prosody becomes very dark indeed: people are accused of orchestrating *fraud, campaigns* to destroy politician’s careers, to “bring down the Twin Towers and the killing of 1.5 million Armenian Christians” (Partington 2017: 195). If native speakers are not necessarily consciously aware of these differences, non-native speakers are likely to experience even more difficulties, particularly if their language of habitual use has cognate words that do not share the same primings for their evaluative prosody, making misleading transfer, or interference, very probable. We would suggest that the well-known category of *faux amis*, or false friends, should be extended to include “dubious” friends – those we can trust sometimes, but not always.

In the follow up to our study, an exploratory collocate-driven strategy, combined with a corpus search based on results, revealed an unexpectedly rich set of

suggestions for equivalents of *contaminazione* when used in its positive, figurative sense: *mélange, échange, influence, partage, dialogue, rencontre, interaction, inspiration, lien, collaboration, carrefour, relation, alliance, collision, symbiose, va et vient*.

Further research using more carefully constructed corpora (ideally with sub-corpora of different genres), and a more extensive analysis of concordance lines, could be useful in testing these findings. Yet we would argue that the evidence discussed here, though limited, is convincing evidence of the yet under-studied potential of cross-linguistic collocation analysis for shedding light on evaluative prosody across languages.

In conclusion, this study shows the importance of carrying out corpus-assisted studies of domain-specific and cross-linguistic variation in semantic preference and evaluative prosody, to understand and correctly transfer concepts between languages. The results can be used to raise the awareness of learners and translators of differences in use between supposed translation equivalents in two languages, and to inform the drafting of more usage-aware, and thus more reliable, lexicographic entries.

References

- Aston, G. (1999) "Corpus Use and Learning to Translate", *Textus* 12(2): 289-314.
- Benko, V. (2014) "Aranea: Yet Another Family of (Comparable) Web Corpora", in P. Sojka, A. Horák, I. Kopeček and K. Pala (eds.) *Text, Speech and Dialogue. 17th International Conference, TSD 2014, Brno, Czech Republic, September 8-12, 2014. Proceedings. LNCS 8655*, Springer International Publishing Switzerland, 257-264.
- Berber Sardinha, T. (2000) "Semantic Prosodies in English and Portuguese: A Contrastive Study", *Cuadernos de Filología Inglesa* 9(1): 93-110.

Biber, D. (2012) "Register as a Predictor of Linguistic Variation", *Corpus Linguistics and Linguistic Theory* 8(1):9-37.

Davies, A. (2004) "The native Speaker in Applied Linguistics", in A. Davies and C. Elder (eds.) *The Handbook of Applied Linguistics*, Oxford: Blackwell, 431-450.

Gellerstam, M. (1986) "Translationese in Swedish Novels Translated from English", in L. Wollin and H. Lindquist (eds.) *Translation Studies in Scandinavia*, Lund: CWK Gleerup, 88-95.

Hoey, M. (2005) *Lexical Priming: A New Theory of Words and Language*, London: Routledge.

Hunston, S. (2007) "Semantic Prosody Revisited", *International Journal of Corpus Linguistics* 12(2): 249-268.

Oksefjell Ebeling, S. (2013) "Semantic Prosody in a Cross-linguistic Perspective", *Studies in Variation, Contacts and Change in English*, 13, <http://www.helsinki.fi/varieng/series/volumes/13/ebeling/>.

Partington, A. (2004) "'Utterly Content in Each Other's Company'. Semantic Prosody and Semantic Preference", *International Journal of Corpus Linguistics* 9 (1): 131-156.

----- (2017) "Evaluative Clash, Evaluative Cohesion and How We Actually Read Evaluation in Texts", *Journal of Pragmatics* 117: 190-203.

----- (2018) "The evaluative Prosody of Forms of Government and Power: -cracy and -archy. Is this Prosody Inheritable and What Are the Implications for Translation?" in *TRALinea*, www.intralinea.org/specials/article/2299.

Sartre, J. P. (1936) *L'Imagination*, Paris: Félix Alcan.

Scott, M. (2020) *WordSmith Tools version 8*, Stroud: Lexical Analysis Software.

Sinclair, J. (1991) *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.

----- (2004) *Trust the Text: Language, Corpus and Discourse*, London: Routledge.

Stewart, D. (2005) "Hidden Culture: Using the British National Corpus with Language Learners to Investigate Collocational Behaviour, Wordplay and Culture-Specific References", in G. Barnbrook, P. Danielsson and M. Mahlberg (eds.) *Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora*, London: Continuum, 83-95.

----- (2010) *Semantic Prosody: A Critical Evaluation*, London: Routledge.

Stubbs, M. (2001) "Computer-assisted Text and Corpus Analysis: Lexical Cohesion and Communicative Competence", in D. Schiffrin, D. Tannen, and H.E. Hamilton (eds.) *The Handbook of Discourse Analysis*, Oxford: Blackwell, 304-320.

Thompson, G. (1996) *Introducing Functional Grammar*, London: Edward Arnold.

Wray, A. (2002) *Formulaic Language and the Lexicon*, Cambridge: Cambridge University Press.

Xiao, R. and A. McEnery (2006) "Collocation, Semantic Prosody and Near Synonymy: A Cross-Linguistic Perspective", *Applied Linguistics* 27(1):102-129.