

Fixing Automatic Metaphor-Literal Alignments

Prisca Piccirilli

February 2022

Contents

1	Annotation Overview	2
2	Metaphors vs. Idioms	2
3	Notation: Column "Keep"	2
3.1	Instances that We Keep	2
3.2	Instances that We Discard	2
3.3	Interesting Cases	2
4	Instance That We Keep: 1	3
4.1	General Rule	3
4.2	Pronouns	3
4.3	Quantity Nouns	3
4.4	Phrasal Verbs	4
4.5	Passives	4
5	Instance We Do Not Keep: 0	4
5.1	Both Components Present But Do Not Belong Together	4
5.2	Verb as Noun - NN expression	4
5.3	Ellipsis	4
5.4	Different Meaning	5
6	Instance We Do Not Keep But Are Interesting: 2 or I	5
6.1	Special case: Expression Appears as an Idiom	5
6.2	Verb Participle as Adjective	5
6.3	Object as Subject - SV expression	5
6.4	Noun Phrases Containing the Object Component	5
7	Alignment Rules	6
7.1	General Rule	6
7.2	Pronouns	6
7.3	Prepositions/Particles	6
7.4	Verb Tenses	7
7.5	"NULL" Translations	8
7.6	More Than One Occurrence Of VO-pair	8
7.7	Different Translations	9

1 Annotation Overview

After extracting aligned parallel sentences in English-French/German containing our synonymous metaphorical vs. literal expressions (such as *tackle question* vs. *address question*), we obtained the word alignments using *fast-align* for the components of the expressions, i.e., what is *tackle* mapped to, what is *question* mapped to, in French and German. However, as both the extraction and the alignments were automatically done, there are some errors in (i) the instances we are interested in, (ii) the alignments of the expressions in English to French/German.

The goal of this annotation project is of two folds:

First, one needs to *keep* or *discard* the instances if they do not respect our "expressions" conditions.

Then, for each *kept* instance, one needs to assure that the alignments for the components of the corresponding expression are *correct* as well.

We give below an explanation with examples of the different cases we have faced so far. Keep in mind that this is a work in progress, therefore the guidelines can be slightly modified and further examples and difficult cases might be still missing.

2 Metaphors vs. Idioms

This work is about *metaphors*, not to be confused with *idioms*. Both fall under the realm of *figurative language*, but it is important to distinguish them. As you will see, some extracted instances will contain *idioms* instead of *metaphors*, which you will annotate as 2 (Section 6.1.)

Idioms, or *idiomatic expressions*, are non-compositional phrases whose meaning cannot be derived from the meaning of their constituents, e.g., *raining cats and dogs*, *under the weather*, *break the ice*, *bark up the wrong tree*, *spill the beans*, etc. They tend to be *fixed expressions* that are lexically not flexible, i.e., applying some small variation to an expression, as in **spill the peas*, **spill some beans*, **spill the black beans*, will not preserve the idiomatic meaning.

Metaphors are a bit more subtle than *idioms* and are defined as "understanding and experiencing something in terms of another" (Lakoff and Johnson, 2003). That "something" is usually ABSTRACT, but expressed in a CONCRETE way. In the expression *grasp the meaning*, the abstract concept of "understanding" is expressed through the concrete concept of "grasping" (i.e., seize, hold), which one can bodily mimic.

Contrarily to *idioms*, which can be interpreted both literally and idiomatically, depending on the context, e.g., "Dad had to *break the ice* on the chicken troughs so that they could get water" vs. "Next week the two diplomats will meet in an attempt to *break the ice* between the two nations", *metaphors* can only be interpreted *metaphorically* as they violate selectional preference (Wilks, 1971) (one cannot *literally* "grasp" a "meaning").

3 Notation: Column "Keep"

In these guidelines, all examples in **blue** are examples of instances that we *keep* whereas examples in **red** are examples of instances that we discard. Examples in **orange** are instances that we neither keep or discard as they do not "respect" certain conditions but are interesting cases nonetheless.

3.1 Instances that We Keep

- Notation: 1

3.2 Instances that We Discard

- Notation: 0

3.3 Interesting Cases

- Notation: 2 or I (capital "i")

4 Instance That We Keep: 1

4.1 General Rule

Our expressions are of VERB-OBJECT (VO) form, as listed in Table 7.7. In the first and second column (*pair* and *inflected_pair*, respectively) of the file, it is specified which expression is targeted in that instance.

RULE: If the components (verb + object) of the expression occurs in the English sentence under the form and only the form VO, where the object is modifying the verb that sentence and its translation needs to be kept.

- ('tackle', 'question') ('tackle', 'question') I should like to make clear that my group is going to propose an amendment deleting the detail of how the Commission should tackle this question .
- ('boost', 'economy') ('boosting', 'economy') It has instead resulted in steady progress in terms of democracy and in your commitment , and the commitment of your country , to developing good neighbourly relations , respecting the rule of law , boosting the economy through trade and cooperation and working through the Ohrid Framework Agreement over time .

Sometimes, one has to play around with reformulating the structure of the sentence to see that the object is indeed modifying the verb, but occurs as OV in the sentence. These sentences are kept.

- ('find', 'excuse') ('found', 'excuse') It does not matter who you are or what excuse you think you have found for your perverse and nihilistic , violent ideology . "You think you have *found an excuse*" (VO-pair correct)

4.2 Pronouns

RULE: if a pronoun is referring and only referring to the object (noun) of the expression, the sentence is kept.

- ('flood', 'market') ('flooded', 'market') It is a sensible solution allowing Member States to monitor and control the market and make sure it is not flooded .
- ('tackle', 'question') ('tackle', 'questions') But there are also questions of society which go very deep and which we have to tackle with the help of women 's [...]
- ('address', 'problem') (('address', 'anzugehen') It is incumbent on the European Union and the United Nations to address what is a social and fatal problem in a comprehensive and strategic manner . "It is incumbent [...] to address the social and fatal problem [...]"

4.3 Quantity Nouns

RULE: if the noun of the expression is first modifying a quantity noun which is modifying the verb of the expression, the sentence is kept.

- ('pour', 'money') ('pour', 'money') The UK will continue to pour increasing amounts of money into Brussels , yet the one area where some of that money is recouped [...]
- ('cast', 'doubt') ('cast', 'doubt') In the UK the previous Tory government cast a great deal of doubt on whether or not the euro would actually happen .
- ('attack', 'problem') ('attacking', 'problem') Ethiopia , a few years ago , was , hopefully , the last case where we were more or less hijacked into pouring in wheat instead of attacking the roots of the problem . "the roots of" is not a quantity noun, "the roots" is directly modifying the verb and the sentence is therefore *discarded* - we mark these examples as 2 (see Section 6.4).

4.4 Phrasal Verbs

RULE: if the verb is encountered in the form of a phrasal verb, but the meaning of the overall expression does not change, the sentence is kept.

- ('pour', 'money') ('pour', 'money') I want the Commissioner to ensure that whatever is brought in is unified throughout the European Union and is financed in a unified way so that no one Member State can **pour in more money** than another to support their particular industry .
- ('suck', 'worker') ('suck', 'workers') We are therefore trying to **suck in the health workers** from other countries in order to deal with our own problems .

4.5 Passives

- ('address', 'question') ('addressed', 'question') At any rate not such typology has been conveyed to us by way of complaints , unlike in the case of DVD , which is the subject on which we are conducting the inquiry and to which your **question is addressed** .

5 Instance We Do Not Keep: 0

5.1 Both Components Present But Do Not Belong Together

RULE: the way English sentences were extracted might result in both components of the target expression being present in the sentence, however if they do not belong together, the sentence is discarded

- ('tackle', 'question') ('tackle', 'question') Rather than go back over the areas that have already been covered I would like to deal with one specific topic : the **question** of drugs and how we **tackle the pervasiveness of the drug culture** in our societies .
- ('attack', 'problem') ('attacking', 'problem') I personally believe that it is nothing short of scandalous that we hear Members of this Parliament **attacking** the demonstration rather than looking at the solution to the **problem** , which is fair conditions , fair pay and fair hours .

5.2 Verb as Noun - NN expression

RULE: both components are present in the sentence and they belong together, but the verb from the VO expression is a noun which results in a NN expression. The sentence is discarded.

- ('boost', 'economy') ('boost', 'economy') As he said , fish farming is a critically significant source of employment in remote parts of rural Scotland and at a time when the rural areas of Scotland are suffering their worst recession in over a hundred years , the aquaculture industry provides an essential **boost to the rural economy** on which many thousands of direct and indirect jobs depend .
- ('abuse', 'alcohol') ('abuse', 'alcohol') Right across Europe , the **abuse of alcohol** by children and adolescents is a growing problem .

5.3 Ellipsis

- ('read', 'book') ('read', 'book') Mr President , thank you for giving me the floor , despite the fact that I have not yet written a **book** and have not **read** a great many .

5.4 Different Meaning

RULE: the meaning of the pair is not the intended one: as a test, replace the verb by its synonymous (metaphorical or literal), if the meaning is different, then the instance is discarded.

- ('address', 'question') ('addressed', 'question') The second question is really more broadly addressed to Parliament : are we going to set up a permanent standing liaison committee of MEPs involved with development and cooperation to report to the General Assembly of the United Nations each year on our joint cooperation ? *The second question is really more broadly *tackled* to Parliament [...]

6 Instance We Do Not Keep But Are Interesting: 2 or I

6.1 Special case: Expression Appears as an Idiom

RULE: the expression appears to be used in an idiomatic way in that context. The sentence is discarded. We mark these instances as *I* (capital "i").

- ('pour', 'money') ('pouring', 'money') I have always maintained that good governance was more important than pure democracy if you are going to stamp out corrupt practices , and if we can not wipe these practices out we are only pouring good money after bad .
- ('make', 'remark') ('made', 'remarks') So the honourable Member 's remarks about the strategic importance of these regions are extremely well made .

6.2 Verb Participle as Adjective

RULE: the participle of the verb is used and acts as an adjective modifying the noun of the expression. The sentence is discarded.

- ('break', 'agreement') ('broken', 'agreement') Unless our leaders agree on a revised and workable framework for fiscal stability , and inscribe this in our basic law , then we are left with nothing but a broken gentlemen 's agreement .

6.3 Object as Subject - SV expression

RULE: both components are present in the sentence and they belong together, but the noun from the VO expression is a subject which results in a SV expression. The sentence is discarded.

- ('fuel', 'debate') ('fuelled', 'debate') It is also regrettable that this debate has taken such a hysterical tone in certain parts of Europe , fuelled by some of the London-based media in particular .
- ('drop', 'price') ('dropped', 'price') For example , in Malta consumers and businesses saw an exorbitant increase in their energy bills when the price of oil was at its high , but there was no decrease even when the price of oil dropped more than half .

6.4 Noun Phrases Containing the Object Component

RULE: the noun XX from a noun phrase of the form "the XX of + object from VO pair", and not the object, is directly modifying the verb from VO pair. The sentence is discarded. Also see Section 4.3.

- ('address', 'problem') ('addressing', 'problem') As he so rightly highlighted , that is the bulk of the problem in terms of the surplus that this Parliament is addressing . It is "the bulk" that is addressed, not "the problem".

7 Alignment Rules

IMPORTANT: Only correct the alignments for the sentences that we keep (in blue). For the others, just apply the correction notation under the column "keep" and continue. The automatic extracted alignments are provided under the 3rd column *alignments*. Alignments are of data type *list of tuples* for each component of the expression. Please keep the same format when proceeding to corrections.

7.1 General Rule

Align **all components** of the source expression (**VERB + OBJECT**) to the **minimal but all** translations in the target language. **Determiners** in the target language ("la", "ce", "der", "diese" etc.) are usually **not included** in the alignments, **unless** they are part of a joint preposition ("au", "aufs") (see section 7.3).

- EN2DE ('colour', 'judgement') ('colouring', 'judgement') [(['colouring', 'getrückt'), ('colouring', 'sein'), ('judgement', 'Urteilsvermögen')]
We should see this Commission as a college and while the individual commissioners must be held accountable , nevertheless we must take an objective view of what is going on and beware of either European party politics or international politics **colouring our judgement** . Wir sollten die Kommission als Kollegium betrachten , und obwohl für die Kommissionsmitglieder das Prinzip der Einzelverantwortung gilt , müssen wir stets objektiv bleiben , und unser **Urteilsvermögen** darf nicht von europäischer Parteipolitik oder internationaler Politik **getrückt sein** .
- EN2DE ('find', 'excuse') ('finding', 'excuse') [(['finding', 'bieten'), ('excuse', 'Entschuldigung')]
Can we really restore trust and generate inspiration among the voters if we continue to conduct our policies from one national election to another , **finding in them a good excuse** not to apply the common policies ? Können wir das Vertrauen wirklich wiederherstellen und die Wähler inspirieren , wenn unsere Politik nur von einer nationalen Wahl zur nächsten reicht und wenn diese Wahlen eine willkommene **Entschuldigung dafür bieten** , dass wir unsere gemeinsamen politischen Strategien nicht umsetzen ?
- EN2FR ('tackle', 'question') ('tackle', 'question') [(['tackle', 's'), ('tackle', 'attaquer'), ('tackle', 'à'), ('question', 'problème')]
I should like to make clear that my group is going to propose an amendment deleting the detail of how the Commission should **tackle this question** . Je voudrais dire clairement que mon groupe va proposer un amendement visant à éliminer les détails concernant la manière dont la Commission devrait **s'attaquer à ce problème**. Note that the determinant "ce" is not included in the alignments.

7.2 Pronouns

RULE: If the noun from the English VO pair is replaced as a pronoun in the target language, map the English noun to the corresponding noun the target pronoun is referring to .

- EN2FR ('address', 'problem') ('address', 'problem') [(['problem', 'problème'), ('address', 'régler')]
The first point is the importance of this issue being treated on a bottom-up basis rather than a top-down basis , that it is up to Member States as close as possible to the problem to **address the problem** most successfully . Premier point : l'importance de traiter la problématique de l'exclusion sociale d'un point de vue ascendant plutôt que descendant ; ce sont les États membres les plus proches du **problème** qui sont les mieux à même de **le régler** .

7.3 Prepositions/Particles

RULE 1: Map only "as the translation was produced", or in other words, what you can "copy paste". Do not separate prepositions with their articles when they are joint.

- EN2FR ('attack', 'problem') ('attacking', 'problem') [(['attacking', 's'), ('attacking', 'attaquer'), ('attacking', 'au'), ('problem', 'problème')]
Moreover , as poverty and

lack of local opportunities are amongst the main reasons for children being trafficked , the poverty-reduction approach of our development cooperation is a meaningful basis for attacking the problem .

De plus , comme la pauvreté et l' absence d' opportunités locales sont parmi les principales causes du trafic d' enfants , l' approche visant à réduire la pauvreté de notre coopération au développement constitue une excellente base pour s' attaquer au problème ."

RULE 2: Include the preposition in the alignments when it is "part of" (particle) the verb/noun meaning in the target language.

- **EN2FR** ('break', 'agreement') ('break', 'agreement') [('break', 'rupture'), ('break', 'de'), ('Agreement', 'accord')] Right now , if we break the Interinstitutional Agreement , we go back to the Treaty , which gives Parliament a lot more power and a lot more authority . À présent , en cas de rupture de l' accord interinstitutionnel , nous en revenons au Traité , qui confère au Parlement bien davantage de pouvoir et d' autorité ."
- **EN2DE** ('tackle', 'question') ('tackle', 'question') [('tackle', 'in'), ('tackle', 'Angriff'), ('tackle', 'nehmen'), ('question', 'Frage')] However , one needs to underline that if we really want to tackle the question of security on board aircraft then cost considerations can only be secondary . Doch man muss unterstreichen , dass Kostenerwägungen nur zweitrangig sein können , wenn wir die Frage der Sicherheit an Bord von Flugzeugen wirklich in Angriff nehmen wollen . Note that the determinant "die" is not included in the alignments.
- **NOTE:** Rule 2 also applies for present participles, e.g., *attacking the problem* if translated into *en s'attaquant au problème*, "en" is part of the verb as this is how present participle is built in French, so the alignment would be [('attacking', 'en'), ('attacking', "s"), ('attacking', 'attaquant'), ('attacking', 'au'), ('problem', 'problème')]

7.4 Verb Tenses

RULE: Align all components of verb tenses to keep track of kept or changed tenses between the source and target language.

- **EN2FR** ('tackle', 'question') ('tackle', 'questions') [('will', 'traitera'), ('tackle', 'traitera'), ('questions', 'questions')] A strategy paper will draw on the results of this debate and tackle fundamental questions about the future of Europe . Une étude stratégique mettra à profit les résultats de ce débat et traitera les questions fondamentales concernant l' avenir de l' Europe .
- **EN2DE** ('tackle', 'question') ('tackle', 'questions') [('tackle', 'wird'), ('tackle', 'befassen'), ('questions', 'Fragen')] A strategy paper will draw on the results of this debate and tackle fundamental questions about the future of Europe . Ein Strategiepapier wird sich ausgehend von den Ergebnissen dieser Debatte mit den grundlegenden Fragen zur Zukunft Europas befassen .
- **EN2FR** ('float', 'idea') ('floated', 'ideas') [('ideas', 'idées'), ('have', 'ont'), ('been', 'été'), ('floated', 'avancées')] Many ideas have been floated on a comprehensive approach to peace in the Middle East . De nombreuses idées ont été avancées quant à une approche globale de la paix au Moyen-Orient .
- **EN2FR** ('address', 'problem') ('addressed', 'problems') [('problems', 'problèmes'), ('are', 'sont'), ('being', 'sont'), ('addressed', 'traités')] Indeed , rail pricing problems are being addressed by us in the framework of the follow-up to our Commission White Paper on railways strategy which we published last year . En effet , les problèmes liés aux tarifs ferroviaires sont traités dans le cadre du suivi du livre blanc de la Commission sur la stratégie en matière de transport ferroviaire publié l' année dernière .

- **EN2DE** ('address', 'problem') ('addressed', 'problems') [('problems', 'probleme'), ('are', 'werden'), ('being', 'werden'), ('addressed', 'behandelt')] Indeed , rail pricing problems are being addressed by us in the framework of the follow-up to our Commission White Paper on railways strategy which we published last year . Bahntarifprobleme werden ja auch von uns im Rahmen des Follow-up des im letzten Jahr veröffentlichten Weißbuchs der Kommission über Eisenbahnstrategie behandelt .
- **EN2FR** ('breathe', 'life') ('breathing', 'life') [('are', 'sommes'), ('breathing', 'en'), ('breathing', 'train'), ('breathing', 'de'), ('breathing', 'redonner'), ('life', 'vie')] We are breathing new life into the Western European Union , which should have actually been defunct last year . Or , nous sommes en train de redonner vie à l' Union de l' Europe occidentale , qui aurait dû rendre son dernier souffle l' an dernier .”

7.5 "NULL" Translations

RULE: If one of the expression's component is omitted in the translation, map that component to "NULL".

- **EN2FR** ('tackle', 'question') ('tackling', 'questions') [('tackling', 'NULL'), ('question', 'questions')] Concerning the question to the Commission on the implementation of the EU-Russia visa facilitation agreement that I tabled on behalf of the ALDE Group , I wanted to learn what progress has been made so far and whether we are , at the moment , in a position to expect a breakthrough in tackling technical questions such as the requirement for EU citizens to register with the authorities within three days if staying in a private home in Russia . Concernant la question sur la mise en œuvre de l' accord visant à faciliter la délivrance de visas entre l' Union européenne et la Russie que j' ai posée à la Commission au nom du groupe ALDE , je souhaitais savoir quels progrès avaient été réalisés jusqu'à présent et si nous pouvions , actuellement , nous attendre à une avancée sur les questions techniques telles que l' obligation pour les citoyens de l' UE de s' enregistrer auprès des autorités dans les trois jours si leur séjour se déroule chez un particulier en Russie .”

7.6 More Than One Occurrence Of VO-pair

RULE: If the VO pair occurs more than once in the English sentence, repeat the sentence and its translation as many times as the VO pair occurrence, to provide all translations accordingly.

- **EN2FR** ('address', 'problem') ('address', 'problem') [('address', 'nous'), ('address', 'attaquons'), ('address', 'au'), ('problem', 'problème')] Unless we address the very serious problem of the declining age profile in European agriculture , in the not too distant future Parliament will be sitting in an emergency session to address what has turned out to be a very serious problem . Si nous ne nous attaquons pas au problème très grave du vieillissement des agriculteurs européens , le Parlement siégera bientôt en séance d' urgence pour faire face à un problème qui est devenu très sérieux .

('address', 'problem') ('address', 'problem') [('address', 'faire'), ('address', 'face'), ('address', 'à'), ('problem', 'problème')] Unless we address the very serious problem of the declining age profile in European agriculture , in the not too distant future Parliament will be sitting in an emergency session to address what has turned out to be a very serious problem . Si nous ne nous attaquons pas au problème très grave du vieillissement des agriculteurs européens , le Parlement siégera bientôt en séance d' urgence pour faire face à un problème qui est devenu très sérieux .

- **EN2FR** ('address', 'problem') ('addressing', 'problems') [('addressing', 'en'), ('addressing', 's'), ('addressing', 'attaquant'), ('addressing', 'aux'), ('problems', 'problèmes')] A good start has been made in addressing the structural problems facing Georgia , tackling for example endemic corruption , which has harmed every facet of life in Georgia . Un bon départ a été donné en s' attaquant aux problèmes structurels du pays , tels que la corruption endémique , qui a eu des influences néfastes sur tous les aspects de la vie en Géorgie .

7.7 Different Translations

RULE: If the VO pair in English is translated by a different VO pair in the target language, align accordingly.

- **EN2FR** ('make', 'remark') ('make', 'remarks') [('make', 'donner'), ('remarks', 'conclusion')]" Two questions have been put very clearly and I would like to answer them before I make my final remarks . Deux questions ont été posées très clairement et je vais y répondre avant de donner ma conclusion .

Original VO Pairs	Extended Arguments
throw-make remark	comment
break-end agreement	cycle, relationship, contract, marriage, process,pattern
cast-cause doubt	issue, fear
grasp-understand meaning	concept, issue, point, problem, situation, reason, language, idea, risk, question
tackle-address question	issue, problem, concern, challenge, situation, point, crisis, matter, inequality, task
catch-get disease	idea, chance, information, result, message, point, call, problem, opportunity
leak-disclose report	information, document, story
shape-determine result	life, outcome, success, strategy
stir-cause excitement	confusion, reaction, feeling, emotion
follow-practice profession	religion, activity
disown-reject past	idea, policy, responsibility
frame-pose question	problem, challenge, issue, debate, concern, argument, idea, hypothesis
boost-improve economy	service, system, situation, process, education, work, business, result, number
abuse-misuse alcohol	drug, substance, medication, product
pour-invest money	fortune
wear-have smile	X
devour-read book	article, story, page, novel, information, chapter, news
attack-address problem	issue, need, question, challenge, point, situation, topic, change, cause
juggle-manage job	project, work, life, career, school
flood-saturate market	X
push-sell drug	X
suck-attract worker	talent
drown-forget trouble	pain, problem, feeling
twist-misinterpret word	fact, meaning, comment, situation, information, message
poison-corrupt mind	system, process, soul, relationship
fuel-stimulate debate	growth, economy, interest, discussion, demand, activity, imagination, creativity, conversation
taste-experience freedom	pain, life, joy
mount-organize production	event, campaign, conference, exhibition, demonstration, protest
buy-believe story	word, lie
sow-cause doubt	death, confusion, chaos, conflict, panic, fear, violence, uncertainty, terror, hatred
float-suggest idea	theory, concept
dull-decrease appetite	pain, sense, noise, feeling
color-affect judgement	decision, choice, perception, experience, interpretation
breathe-instill life	sense, confidence, value, spirit, love, hope, idea, passion
recapture-recall feeling	memory, moment, experience
cloud-impair memory	ability, judgement, judgment, mind, vision, thinking, perception, understanding
shipwreck-ruin career	X
deflate-reduce economy	cost, price, value, supply, wage, market, currency

absorb-assimilate knowledge	information, idea, culture, material, lesson, fact, content, experience, concept, thought
absorb-pay costs	fee, bill, tax, debt, interest, expense
digest-comprehend information	meaning, material, fact, text, concept, idea, word, content, situation, message
find-make excuse	way, connection
kill-cancel proposal	project, bill, program, process, agreement, deal
close-finalize deal	case, plan, arrangement, agreement, project
close-end investigation	season, deal, case, debate, operation, story
shake-damage confidence	foundation
drop-reduce price	cost, rate, temperature

List of original metaphorical-literal VO pairs and their extended arguments.