

# How to Translate Multi-Language E-Discovery Search Terms

Bryan Quinn, Director of Business Development, TLS Oslo



If you read our previous blog on multi-language e-discovery, you're aware of the challenges faced when dealing with international matters. Searching across languages with the same degree of accuracy and proportionality can be equal parts art and science. Different languages, by their nature, act in different ways.

In the haste to get document review started, the nuances to effectuating a proper multi-language search strategy can often be overlooked. With TransPerfect being one of the world's largest translation providers, and TLS being an e-discovery powerhouse, this is an area where our two worlds most evidently collide.

In this piece, we outline a few translation and grammar concerns to consider when drafting your multi-language terms.

## Synonyms

A successful search strategy requires a high level of awareness of potential synonyms, along with an understanding of their prevalence and interchangeability in other languages.

Best-selling author Elizabeth Day frequently writes on the topic of failure; on a promotional tour in Amsterdam, she learnt that Dutch has two words for failure.

'Falen' is used for common failures such as exam results, whereas 'pech' is used for failures beyond a person's control, those of 'existential bad luck'. It's easy to see that the difference between 'falen' and 'pech' can have dramatic legal differences and implications.



@TransPerfectLegalSolutions



@TLSlegal

## Localisation or Localization?

If drafting terms in an automobile case, would you search 'hood' or 'bonnet', 'trunk' or 'boot', 'windshield' or 'windscreen'?

Native-level speakers of a language are normally well-versed at identifying its regional differences. But being cognisant (or cognizant) of these differences in other languages is equally important.

Text search does not forgive using the wrong regional variants. Improperly localised terms mean the case's keyword list is either 'cocked up' or 'screwed up' depending on which side of the Atlantic you're on.

## Slang, Euphemisms, and Idioms

Slang terms are often the most important keywords in a case. They're also one of the most frequently mistranslated terms. The risk comes from using a literal translation, and the biggest culprits of literal translations are free, online machine translation tools.

'Grease the palm' directly translated into Portuguese is nonsense. The proper translation is 'molhar as mãos', which literally translated back to English is 'wet hands'. To the point on synonyms, 'dar luvas', 'por baixo da mesa' and 'untar as mãos' are also commonly used to express the same concept.

Slang is also highly localised, so translators should be drawn from the region of the data and given the full context.

Portuguese-speaking procurement managers in Sao Paulo, Angola and the Algarve will use wholly different languages when soliciting bribes.

## Names

Europeans are often typing on keyboards that don't have their vowels, so they substitute. 'Chloé' can frequently appear as 'Chloe' but from a search perspective these are two different words.

Moreover, we're often dealing with different alphabets.

Not only do the Russian and English alphabets have different characters, but the English alphabet has 26 letters whereas the Russian has 33. Russian speakers communicating with the West can substitute English characters for the Cyrillic to fill the gaps.

A best practice for searching the name Alex (short for Alexander) in a data set containing Russian and English may require adding the English terms 'Sasha', 'Aleksandr', 'Aleksander', 'Sashenka', 'Sanya', 'Sashka', 'Sanyok' and 'Aleks', as well as the equivalents 'Саша', 'Алекс', 'Александр', 'Сашка', 'Сашенька', 'Саня', or 'Санёк' in Russian.

We also know companies work hard to stay on brand across the global marketplace, but even the world's largest can have idiosyncrasies when translating their names for international markets.

Coca-Cola is the second most widely understood phrase in the world (after 'ok'), and despite efforts to keep a single spelling and logo, the brand name is still transliterated into at least 15 languages.



@TransPerfectLegalSolutions



@TSLegal

## Gender

If you've ever studied a language, you know assigning different genders to nouns is common.

Perhaps you learnt this fun fact from Netflix's recent smash hit "Emily in Paris", where the protagonist is shocked to discover that parts of the female anatomy are qualified as masculine ("le" instead of "la").

However, there are almost as many exceptions as there are rules.

Gender designation of nouns is significant for many reasons. It can affect the spelling of adjectives and whether they must agree. To admit we're both old in French, is to say Rob is 'vieux' whereas Alys is 'vieille'. Furthermore, an analysis would be 'indépendante' but an audit would be 'indépendant'. To search just one spelling iteration would miss the other.

## Plurals

Many Asian languages don't have plurals as we understand in English. To make matters more interesting, some languages have more than one way to form a plural and, in any language, extra attention is owed to irregular nouns (such as one foot but two feet).

In Thai, you can add a numerical designator after the noun so 'five cats' becomes 'cat five', or you can simply repeat an object so 'two girls' becomes 'girl girl' (although that could also mean multiple girls!).

If plurals are included in a list of keywords to translate but the number and context aren't given, like synonyms, the different options will neither be understood nor provided.

## Wildcards and Stem Searching

Following from the last example, all types of wildcard searches are a minefield, and we're often asked to translate truncated terms. Our advice is to always exercise great caution.

Root words are not structured the same way across languages. Moreover, roots frequently do not exist in the same form and, when they do, they won't capture all the same variations – or they may capture too many.

Arabic is a prime example of a language founded on root terms, and the Arabic dictionary is organised alphabetically by root word. As such, the same three-to-six letter root could hit on the terms 'book', 'books' and 'he wrote', but also 'officer' and 'library'. Arabic also reads right to left, so careful consideration must be taken on where to place the wildcard.

How we handle wildcards is done on a term-by-term basis and is heavily impacted by the language. One of many strategies is defining the full list of term variants, getting a translation for each and working backwards from there.

## Word Order

Different languages order nouns, adjectives and numerical designators in different orders, and searching for specific phrases can therefore be impacted.

Continuing the Thai example, 'six red apples' is 'apple red six', and 'five green dollars' can be translated as 'five dollar colour green' because the designator itself is commonly added.

This is a particularly notable consideration when revising terms after the initial hit report. Simply adding or removing words can upset natural language usage thereby rendering a term translation null and void.

## False Friends

Just because a word exists in two languages doesn't mean its meaning and usage are the same. A French general counsel may well speak to an English lawyer about a 'sensible' commercial decision without the lawyer realising it's in fact a 'sensitive' commercial decision.

## Homonyms

It's important to remember that many words have multiple meanings. With keywords, it's vital that translators are given context, or some embarrassing blunders can and will occur.

One can close a deal, be physically close, be intimately close, close a box, and, in the UK, also live on a Close. In the vast majority of languages, each 'close' will have a different translation. If working without the context, a translator isn't wrong giving any translated variant, but the net effect could be dramatically different: all the relevant material to 'the closed deal' being omitted with all communications about extramarital dalliances being included.

## Proximity Searches

Finally, there are many syntactical considerations for finalising multi-language terms, but we'll leave you with one of the most important: some languages naturally have more or less words than others.

To yield the same results for the English term 'TransPerfect W/10 ediscovery', the proximity operator should reflect W/13 words in Spanish, W/15 in Hebrew, and W/8 for German.

