# TERMPROFILE.COM - SUPPORTING A CONFERENCE INTERPRETER'S WORKFLOW

Anja Ruetten
(ruetten@sprachmanagement.net)

freelance conference interpreter
Düsseldorf, Germany
www.sprachmanagement.net
www.termprofile.com

## ABSTRACT

Termprofile.com is a tool that helps you to search for terms on the Internet, distinguishing between the countries of origin of the hits, presenting relative frequency and comparative numbers for alternative terms in one language or different languages at a glance, TermProfile is one of the instruments that can significantly enhance interpreters' (and of course translators') workflow. TermProfile shows whether terms or expressions in a certain language are used in the respective countries. It offers an interface to perform three queries in parallel and then displays the search result figures on one screen. For each query it is possible to enter a second term ("control term") in order to obtain a relative frequency of a term. TermProfile offers a web-statistics profile of terms and expressions with dimensions and relations that far exceed a simple Google search.

## 1. THE SITUATION

The Internet is becoming increasingly popular as an instrument for checking the commonness of terms. Translators, interpreters, journalists, technical editors, terminologists and authors all check the frequency of use of terms by entering them in search engines such as Google.

The Internet is a huge text corpus of all languages, countries, topics, population groups - free-of-charge and in real time. An opportunity. Search engines generally provide hits of all possible origins; there is no linguistic (and certainly not an editorial) check. A risk. This is because seemingly common terms can be based on faulty translations and thus be fairly common.

## 2. THE TERMPROFILE PRINCIPLE

TermProfile shows whether terms are also used in countries where the native languages are spoken. In addition, TermProfile provides an at-a-glance list of figures. In other words, a **web-statistics profile** of the checked expressions with comparative figures for alternative expressions or various countries of origin.

TermProfile presents statistic correlations and orders of magnitude about the use of expressions in countries and language areas, which are not visible with a simple entry in Google. Of course, you can analyse terms both in the narrower sense of "specialist terms", but also as general expressions, sayings, theoretically any possible character sequence.

Not least, TermProfile trains your sense of dimensions of the frequency of use of expressions by language, country and specialist subject.

## 3. THE TECHNICAL SIDE

TermProfile is based on an official interface to the data set of Google. Via a so-termed API, short for Application Programming Interface, TermProfile forwards complex queries to the Google database. A part of the information on the results is the statement on the country of origin of the respective web document.

Accordingly, the search functions like those in Google also apply, e.g. the use of quotation marks for groups of words and Boolean operators like the automatic "AND" search, the use of "OR" for alternative terms, etc.

TermProfile is in its experimental phase. At this stage, Google enables 1,000 queries a day. Therefore, the availability of TermProfile is not guaranteed. Technical comments and suggestions are highly appreciated.

## 4. TERMPROFILE IN DETAIL

### 4.1 CORRECTNESS IS RELATIVE - PAY ATTENTION TO THE COUNTRY OF ORIGIN

#### THE FUNCTION

You can **limit** the **search** of a term **to certain countries** (or in the case of EN and ES, country groups). By way of comparison, you can perform a search for this term **in parallel in several countries/country groups and worldwide.** The hits for the various country constellations are presented **as ratios.**

#### WHAT IT MEANS

- How frequently is a term used in one country in relation to other countries?
- How frequent is a second or third term in comparison to this?
One indicator for the correctness of the language used is the country of origin of the text, if we assume that there is a greater probability that a native speaker in the respective country has written the text.

If e.g. a term has disproportionately more hits worldwide than in the countries in which this language is spoken, it is less probable that this term is correct. If the term "muñón del eje" is found on only a few or no Spanish-speaking sites, but on a few (numerous) German sites, this indicates that it might be a translation of the German "Achsstummel" (axle end or axle journal) due to lack of a genuine equivalent in the target language, and that it is not a term coined in Spanish-speaking countries. In this case, further investigation is advisable. If an English collocation in relation to the total hits is rarely found in US, UK, AU, CN, it is less probable that this term is correct. Regionalisms (e.g. all Spanish-speaking countries v Spain; US v UK) can also be recognised in this way.

The number of hits for the respective countries in relation to the number of worldwide hits are displayed in the second results column in order to illustrate the specific ratio. The ratio of hits of "muñón del eje" in Germany to hits worldwide is 1:13; in Spanish-speaking countries 1:11 (thus for one hit in Spanish-speaking countries, there are 11 on the whole Internet). The ratio should actually be more in favour of Spanish-speaking countries (thus the number on the right should be lower). If all hits on the whole Internet could be found in this pre-defined group of Spanish-speaking countries, the ratio would be 1:1.

## Search for 1st token

Search term:     "muñón del eje"

Compare with:

Please select a country for 1st result set:
All major Spanish-/Castillian-speaking countries  ⌄

Please select a country for 2nd result set:
Germany                                            ⌄

Please select a country for 3rd result set:
All major English-speaking countries               ⌄

## Search for 2nd token

Search term:     Achsstummel

Compare with:

Please select a country for 1st result set:
Germany                                            ⌄

Please select a country for 2nd result set:
All major English-speaking countries               ⌄

Please select a country for 3rd result set:
All major English-speaking countries               ⌄

---

### 1st token: "muñón del eje"

| country setting | results for "muñón del eje" | results for country / worldwide results for "muñón del eje" | results for "" | results for "muñón del eje" / results for "" |
|---|---|---|---|---|
| All major Spanish-/Castillian-speaking countries | 11 | 1 : 11.273 | 0 | 1 : 0.000 |
| Germany | 9 | 1 : 13.778 | 0 | 1 : 0.000 |
| total | 20 | 1 : 6.200 | 0 | 0 |
| worldwide | 1 24 | 1 : 1.000 | 0 | 0 |

### 2nd token: Achsstummel

| country setting | results for "Achsstummel" | results for country / worldwide results for "Achsstummel" | results for "" | results for "Achsstummel" / results for "" |
|---|---|---|---|---|
| Germany | 1 900 | 1 : 1.037 | 0 | 1 : 0.000 |
| total | 1 900 | 1 : 1.037 | 0 | 0 |
| worldwide | 1 970 | 1 : 1.000 | 0 | 0 |

"Achsstummel" is 95.000 times more frequent than "muñón del eje" regarding these results

---

### Search results for "muñón del eje" for All major Spanish-/Castillian-speaking countries

Your search for "muñón del eje" produced 1 1

### Search results for Achsstummel for Germany

Your search for Achsstummel produced 1 900 hits.

- anthes :o•n ...Hier finden Sie Schraube

---

## 4.2 FREQUENCY IS RELATIVE – CHECK THE ALTERNATIVES

### THE FUNCTION

By way of comparison, you can perform a search for **several terms in parallel** (and then search for these respectively in the various country constellations). The hits for the various terms and countries are presented as **ratios.**
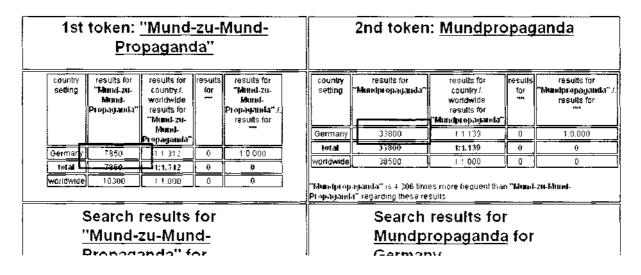
### WHAT IT MEANS

A term may be frequently used in a certain country or a group of countries and still be wrong. That is why it always makes sense to consider whether there are any alternative terms/synonyms in this

language which may be more common or more correct.
Examples:

a) The term "Mund-zu-Mund-Propaganda" appears to have a large number of hits - 78,300; but if you perform a check you will see that "Mundpropaganda" - with 275,000 hits - is ten-times as common.
b) The English equivalent of "Nutzungspopularität" is "user popularity" and not "usage popularity".

## Search for 1st token

Search term: Mund-zu-Mund-Propaganda

Compare with:

Please select a country for 1st result set:
Germany ⌄

Please select a country for 2nd result set:
All major English-speaking countries ⌄

Please select a country for 3rd result set:
All major English-speaking countries ⌄

## Search for 2nd token

Search term: Mundpropaganda

Compare with:

Please select a country for 1st result set:
Germany ⌄

Please select a country for 2nd result set:
All major English-speaking countries ⌄

Please select a country for 3rd result set:
All major English-speaking countries ⌄

### 1st token: "Mund-zu-Mund-Propaganda"

| country setting | results for "Mund-zu-Mund-Propaganda" | results for country / worldwide results for "Mund-zu-Mund-Propaganda" | results for "" | results for "Mund-zu-Mund-Propaganda" / results for "" |
|---|---|---|---|---|
| Germany | 7850 | 1:1.312 | 0 | 1:0.000 |
| total | 7850 | 1:1.312 | 0 | 0 |
| worldwide | 10300 | 1:1.000 | 0 | 0 |

### 2nd token: Mundpropaganda

| country setting | results for "Mundpropaganda" | results for country / worldwide results for "Mundpropaganda" | results for "" | results for "Mundpropaganda" / results for "" |
|---|---|---|---|---|
| Germany | 33800 | 1:1.139 | 0 | 1:0.000 |
| total | 33800 | 1:1.139 | 0 | 0 |
| worldwide | 38500 | 1:1.000 | 0 | 0 |

"Mundpropaganda" is 4.306 times more frequent than "Mund-zu-Mund-Propaganda" regarding these results.

### Search results for "Mund-zu-Mund-Propaganda" for

### Search results for Mundpropaganda for Germany

## 4.3 COMMONNESS IS RELATIVE - PRESENT HITS AS RATIOS

### THE FUNCTION

For a term to be checked, you can enter **a second control term,** which is then searched for in the same selected country constellation. The results of both terms arc presented as **ratios.**

### WHAT IT MEANS

- How frequently is a term used in relation to other words of the same/other language?
In order to see the frequency of a term in relation to the overall text stock of the Internet, you can compare the number of hits to the number of hits of a very general, commonly used word such as "and".
Absolute numbers of hits can sometimes be misleading:
500 hits in Germany may have a different status to hits in the USA.

The fact that the word Spaltrohrmotor (in connection with the word "Pumpe") is less frequent than the little word "and" goes without saying. But what is the exact ratio? In Germany, in terms of absolute measurements, the words "Spaltrohrmotor" and "Pumpe" (in one document) are used precisely 22-times more frequently than in Austria; in relation to the little word "and", the Germany frequency of use is however "only" twice as high (1:680,000 in comparison to 1: 1 270,000, i.e. for 1 x "Spaltrohrmotor" in combination with "pump" in Germany, there are 680,000 x "and", and in Austria 1,270,000 x). The corresponding ratios of the English-speaking countries, the USA and the United Kingdom are fairly homogeneous (approx. 1:550,000). In terms of absolute figures, they are significantly higher, but relatively they are definitely comparable to the German-Austrian values or, in particular, the German value.

These ratios should also contribute to you developing a feel for relative frequencies of use of terms.

## Search for 1st token

Search term: Spaltrohrmotor Pumpe

Compare with: und

Please select a country for 1st result set:
Germany

Please select a country for 2nd result set:
Austria

Please select a country for 3rd result set:
All major English-speaking countries

## Search for 2nd token

Search term: "canned motor" Pump

Compare with: and

Please select a country for 1st result set:
All major English-speaking countries

Please select a country for 2nd result set:
United Kingdom

Please select a country for 3rd result set:
United States

### 1st token: Spaltrohrmotor Pumpe

| country setting | results for "Spaltrohrmotor Pumpe" | results for country /. worldwide results for "Spaltrohrmotor Pumpe" | results for "und" | results for "Spaltrohrmotor Pumpe" / results for "und" |
|---|---|---|---|---|
| Germany | 132 | 1:1 333 | 8980000 | 1 680303 030 |
| Austria | 6 | 1 29 333 | 7620000 | 1 1270000 000 |
| total | 138 | 1:1.275 | 9742000 | 1:705942.029 |
| worldwide | 176 | 1:1 000 | 53400000 | 1 303409 091 |

### 2nd token: "canned motor" Pump

| country setting | results for "canned motor" Pump | results for country /. worldwide results for "canned motor" Pump | results for "and" | results for "canned motor" Pump /. results for "and" |
|---|---|---|---|---|
| All major English-speaking countries | 1680 | 1:1 863 | 9260000000 | 1 551190.476 |
| United Kingdom | 96 | 1 32 604 | 55300000 | 1 576041 667 |
| United States | 1510 | 1:2 073 | 853000000 | 1 564900 662 |
| total | 1680 | 1:1.863 | 9260000000 | 1:551190 476 |
| worldwide | 3130 | 1 1 000 | 11900003000 | 1 376996 805 |

According to these results, "canned motor" Pump is 1.2 174 times more frequent than "Spaltrohrmotor Pumpe"

## 5. THE LIMITS

### 5.1 EQUIVALENCY

The possibilities of using web-statistics methods to find out about the equivalency between two expressions of various languages are relatively limited as equivalency is defined on a semantic level and - in contrast to the country of origin of a document or a frequency of use of a term- is scarcely quantifiable or statistically recordable.

Nevertheless, supplementary information is still possible:

- **A similar web-statistics profile** can confirm equivalency or very different frequencies of use may indicate that there has to be another term.
- **Context words** (in the same search entry field as the term you are actually searching for) help to define the meaning (e.g. search for the word "well" together with "water" to avoid finding "well" as an adverb instead of the noun). Ultimately, a look at several direct application contexts is decisive for determining the equivalency.

## 5.2 RELATIVISING FACTORS

- Various **frequencies of use** can be related to different kinds of industrial, cultural, political, etc. environments.
- It may be the case that there is not yet a target language equivalent for a term and this has thus previously only been coined by translators and is thus also in use.
- As some countries are multilingual, the selection of countries might not enable you to precisely limit your search to native language texts. The "digital" origin of a document does not guarantee anything with regard to its genuine origin either. The TermProfile figures are thus only indications.
- Some expressions have several meanings in one language (EN: well) and are therefore more frequent than the equivalent in the other language.

The **origin** of a site does not provide any definitive information on the origin of the site operator either, but merely expresses a probability - just like the other figures.

The **reliability** of an Internet text and thus the language can ultimately be somewhat guaranteed only by checking the author (private person, public institution, specialist company).

- For every query which sends a request to Google, another server is addressed. Consequently, repeated, identical queries may also lead to deviating results.

## 6. IMPROVEMENT

In the future, there are plans for further possible functions:

- Definition of further country groups for French and German.
- An exclusive country group option, e.g. "all countries except Germany".
- Search for domain endings, e.g. org in order to exclusively find sites with relatively "official" content.
- Definition of individual user profiles (default country groups).

Hopefully, suggestions for further refinement will come up with the application of TermProfile by language professionals, so to make the best possible use of the Internet as a huge multilingual text corpus.