

QUICK AND DIRTY: TERMINOLOGY MANAGEMENT FOR THE IMPATIENT

Naomi Sutcliffe de Moraes

Just Right Communications Ltda.

Abstract: Do you note down every term pair you spent more than 60 seconds researching? If not, you are not taking best advantage of your time and resources. You should never have to research the same term twice. The best terminology tool for you depends on what kind of work you do, so no solution is a best fit for everyone. I will show you how to store your terminology quickly so it won't slow you down, how to convert existing glossaries into a flexible format, and how to access your terminology quickly — whatever your budget.

1. INTRODUCTION

If you are reading this, you are probably dissatisfied with your terminology management system, or berating yourself because you do not even have one. Perhaps you try or tried to note down terminology when working, but found it slowed you down too much. But not noting down terminology so that it can quickly be retrieved slows you down even further! It is like the story of the man who got a job painting the white stripe down the middle of the road. On the first day, he painted further than his colleagues, but on each day his productivity decreased, while theirs did not. Why? Because instead of carrying the can of paint with him, he kept running back to it to wet his paintbrush! Do you run back to the dictionary or do a google search more than once for the same term? I hope this article helps you implement a fast, efficient, inexpensive and easy-to-use system so you will never again waste time looking up the same thing twice.

2. WHO IS THIS ARTICLE FOR?

This article is for translators who do not currently store terminology while working, or those who are not sure if they are using the best method, or are unhappy with their current system. It is also for translators who are having difficulty converting their data from one data format to another, because I will explain how to store your data now so that it can be converted into other formats in the future, and along the way I will provide some conversion tips. If you currently store terms inside your translation environment tool (TEnT), and you are happy with its terminology database, you may want to skip to section 8 on indexing others' glossaries.

Note that this article is NOT about terminology management for people in localization, where consistency within a project is very important. I don't do localization, due to my language combination, so I am not qualified to talk about it.

3. WHAT SHOULD YOU KEEP IN MIND WHEN DECIDING WHAT IS BEST FOR YOU?

No one system is the best for everyone. It depends a lot on how you work, how tech-savvy you are, and how much you want to spend. In this section, I will provide some guidelines on what to keep in mind when evaluating the various methods I describe below.

© Naomi J. Sutcliffe de Moraes

Cite as: MORAES, N. J. Sutcliffe de. *Quick and Dirty: Terminology Management for the Impatient*. In "ATA 51st Annual Conference Proceedings," American Translators Association, Alexandria, VA: 2010.

3.1 Ease of Storing as You Go

If the method you adopt is not easy, you probably will not record your brilliant insights or hard-found terms while translating as you go. It is better to choose the easiest method for you that is also compatible with your workflow, but also compatible with other formats you may want to use in the future (upward compatibility). I once tried recording terms on a piece of paper for a job, and I do not think they ever made it into my database. That piece of paper is probably still floating around my office somewhere.

3.2 Ease of Searching

How easy is it to find the term you want? How many clicks or scrolls? If you use a TEnT with an integrated terminology database, it can be configured to look terms up for you automatically — sort of. This depends a bit on the tool's knowledge of conjugations and inflections. One thing I really like is searching with wild cards, but not all tools will permit this. By wild cards I mean typing autom* will find automatic, automobile, automaton, etc.

3.3 Portability

How often do you switch computers? Perhaps you have a desktop in the office and a laptop at home or for traveling? If you interpret, you may want to have access to your terminology on your mobile phone. If you need portability, take this into consideration when choosing your system. For a mobile phone, text files will probably be the best option.

4. WHAT INFORMATION SHOULD YOU STORE FOR EACH TERM?

If you are just starting out, I recommend you store at least:

- **Source term**

Store the source term as it would appear in a text. For example: "at arms" and not "arms, at", which would be common in a dictionary. It will only cause trouble when converting data. Plus, you will always search your data using a manual or automatic computer search tool, which will not know they are the same thing.

- **Target term**

The primary question here is: should my source/target records be one-to-one or one-to-many? An example of each is in Figure 1. I used to be a one-to-one fan, but now I have discovered the advantages of having some terms in the one-to-many format. Note, though, that I translate only in one direction. If you translate in both directions, stick to one-to-one. I do not recommend the many-to-many format, because if you have two words in your source field and later want to use your database with a TEnT and some other tools, automatic look-up will not work.

If you will be looking up your terms in a database or using an indexing program, one-to-many will probably work just fine, and be a lot quicker to input. If you are storing your terms in your TEnT terminology database (or plan to import them into your TEnT terminology database), you need to decide why — to have a quick reminder of the possible translations, or

to be able to insert the translation directly into the target sentence using a control key sequence? If the latter, you will want to store terms in a one-to-one manner. However, if you have a lot of terms stored this way, and a long sentence, you might end up with too many options showing in your TEnT window. In Figure 2, so many options have been offered in the right pane that it would be quicker and easier to type the word "term" than CTRL+12. Even worse, some tools only let you use the CTRL key for options 1-9. My solution is to use macros (in AutoHotKey) for the most common words and use one-to-many terminology records for the terms with more than two common translations. If in doubt, use the one-to-one format. You can always change later, and changing the other direction (from one-to-many to one-to-one) is labor intensive.

- **Domain**

I use a two to four letter code, such as GEN for general terms, GOVT for government terms, EE for electrical engineering terms, etc. I started out using numbers, but kept having to look them up.

- **Source**

As time passes, I am very glad I wrote down the source for some of the early terms I added to my glossary, because it really helps me to decide if I can trust the translation now. It can also be used to indicate when a term is client-specific.

- **Dialect**

Depending on your language combination, you might not need this. I use two letter codes for English language dialects (US, UK, AU, CA, IR) and similarly for Portuguese language dialects (PT, BR, MZ, CV, etc.). For legal translators, this is *crucial!* A deed in England is not the same as a deed in the United States.

- **Other**

I have one catch-all field for anything else I might want to record. I began my terminology odyssey with about 10 fields, but most of them remained blank so I consolidated them. It just causes trouble when you want to convert your data into another format. I put definitions here, examples, or usage notes.

```

[ENG] report
[POR] divulgar
[DOM] GEN
[SOU] Collins
[REM] in the press
[END]

[ENG] promote; publish; report; release; put out; disclose;
communicate; announce; make public
[POR] divulgar
[DOM] GEN
[SOU] various
[REM]
[END]

```

Figure 1 – One-to-one term record and one-to-many term record for English as target language

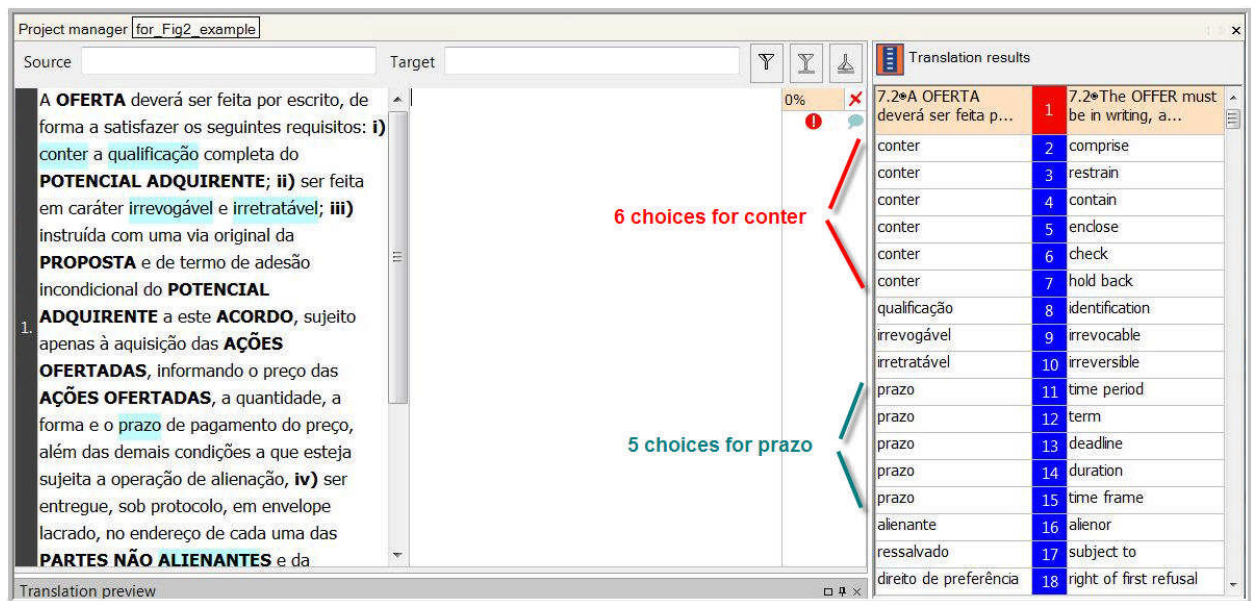


Figure 2 – What can happen in a TEnT when you have a long sentence and use one-to-one term recording

4.1 Tips

Whatever you do, do not use just two columns or fields: one for source and one for everything else. Searching will become a nightmare, and your data will become unconvertible.

I often work with two dialects—European Portuguese and Brazilian Portuguese. I created a separate field for each so I can record the English and both Portuguese terms in the same terminology record.

I have different Word/text files for different areas, but I make them broad areas (law, not criminal law) so I do not have ten files open at once. I use the same codes in the file name as in the terminology records.

For easy convertibility, avoid using commas or tabs in any fields. Use semicolons. Also, avoid automatic capitalization of the first word of a term. Both Word and Excel do this by default, so consider changing the setting. If you later import your terminology into a TEnT, it may cause problems.

5. WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF USING EXCEL?

If you are comfortable with Excel—and I realize not everyone is—Excel is perhaps *the* easiest tool to use to record terminology. I used it for about 5 years. The principal advantage is ease of data entry. You can put all terminology in ONE document, and the document will not become humongous because of how Excel stores data. If you do not intend to add reams of information for each term, this format is great. Just leave the file open while working.

Excel is also great for searching. You can select just one column (source term, usually) and search just in that column, thus avoiding matches in the "other" field. You can also organize your terms into more than one worksheet within Excel, though be warned that you can only search in a single column in one worksheet at a time — if you search on all worksheets, you are also searching in all columns — so don't use too many.

Excel is not good for indexing, which I will discuss in the next section, but it is great for converting to just about any other format. Indeed, you may want to use Excel as an intermediate format for converting data due to its many "Save As" options. The tab-delimited format and the CSV (comma-delimited format) are the two most common formats database and TEnT tools can import. This is why I recommended not using commas or tabs in your terminology entries above.

6. WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF USING WORD OR PLAIN TEXT FILES?

Most people are very comfortable with Word, so data entry is relatively painless. Searching in word files is not so easy, though, without an indexing program. If you put all your terminology files in one folder, you can search for a term using the operating system, but it will only tell you which files contain that word. You must then open the file and search again. Translators do not have time for that! The solution is to use an indexing program, which is great not just for searching your terminology files, but for searching reference files too. I use Archivarius 3000 (www.likasoft.com), which costs only 30 €.

Archivarius is very easy to use. Just pick which folders you want to index, name the index (e.g. terminology), choose the languages in the files to be indexed, and go. Since it understands how words in different languages inflect and searches can be done with wildcards, it is very flexible. In Figure 3, I searched on the word *apresentar*. The left panel shows bits of each file in which the

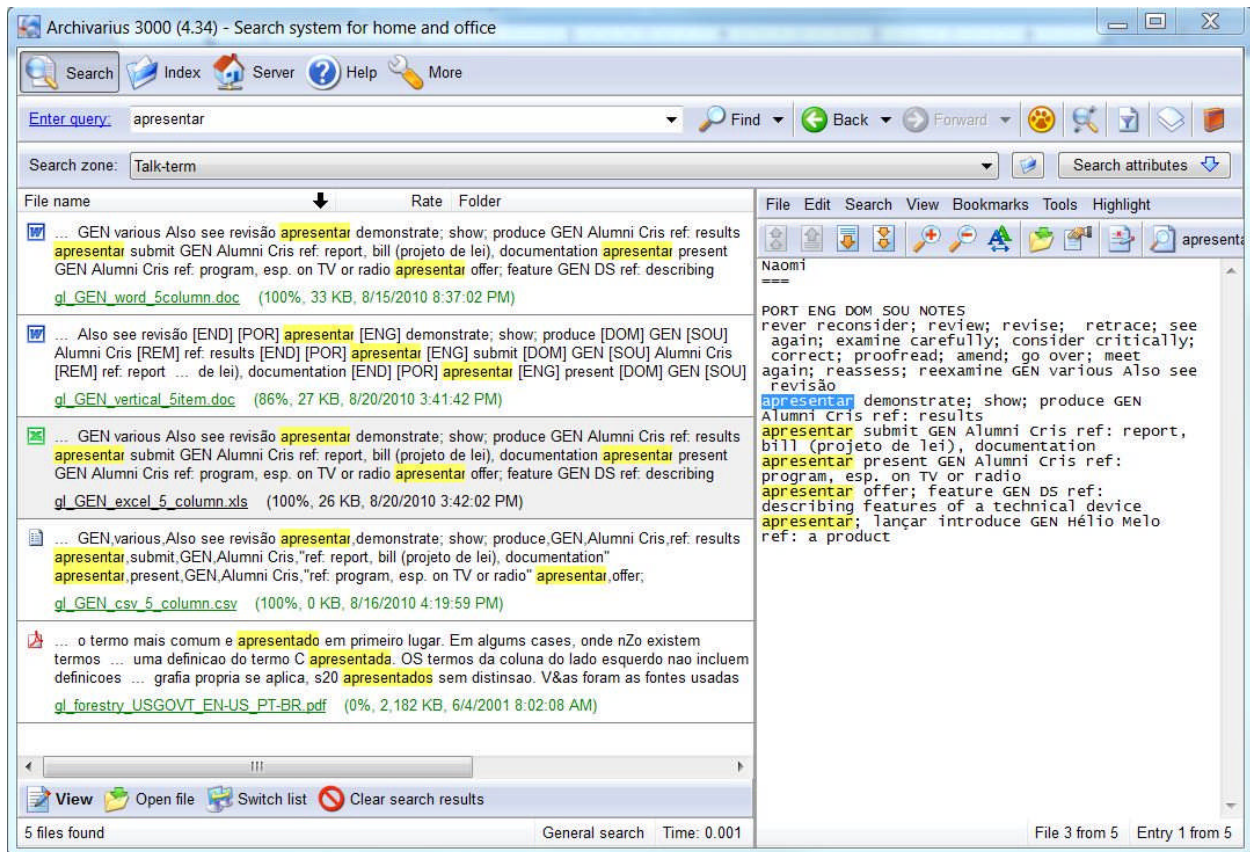


Figure 4 - Search in Archivarius, showing results from Excel terminology records

The drawbacks of using Word files are slow searching, unless you use an indexing program, and large file sizes compared to Excel, though files can easily be converted into simple text files using "Save As" in Word, and become much smaller. Note that Word files with tables are even larger than Word files without tables.

I hope comparing Figures 3 and 4 has convinced you to avoid tables in Word! In order to easily input terms in the vertical format in a Word or text file, I have created a macro in AutoHotKey (www.autohotkey.com) that creates my template when I type ALT+; (see Appendix). Install AutoHotKey, modify my macro for your languages, then click on the file to activate the macro. Put the file in your startup folder to have it active all the time. The program is free and useful for many other tasks as well.

6.1 Converting From the Vertical Format to Other Formats

Terms stored in Excel can be easily converted into tab-delimited or comma-delimited files, as described above. These are the best formats for transfer to other tools, now and in the future. Conversion from the vertical terminology template is easy if you know two simple Word tricks.

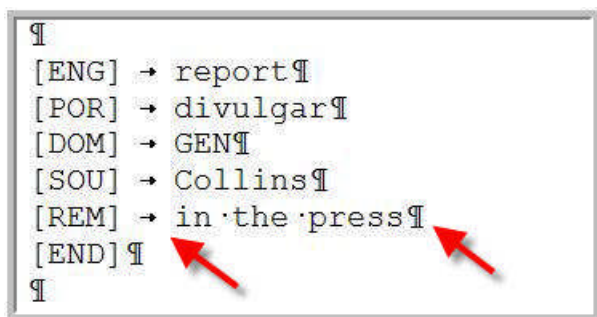


Figure 5 – Terminology template showing formatting

Figure 5 shows the vertical terminology template in Word with formatting displayed. A little-known secret is that you can search on and replace tabs (little arrows) and hard returns (backwards Ps) using the Replace menu option in Word. Tabs are represented by the code ^t and hard returns are represented by the code ^p.

If you have a file with hundreds of terms using the above template, the data can be converted into the tab-delimited format by performing the following replacement functions:

- [ENG]^t replaced by (nothing)
- ^p[POR] replaced by (nothing)
- ^p[DOM] replaced by (nothing)
- ^p[SOU] replaced by (nothing)
- ^p[REM] replaced by (nothing)
- ^p[END]^p replaced by (nothing).

The objective is to remove the tags and have each field separated from the next by a tab. Each record will be on one line, ending with a hard return character. The result of performing these replacements on Figure 5 is shown in Figure 6. Test the replacements on a copy of your file first!



Figure 6 – Terminology template in tab-delimited format, after conversion.

To convert from a tab-delimited or table format to the vertical format, open the file in Excel (or copy the Word table into Excel) and create an extra column before each data field, as shown in Figure 7. Type the code ([ENG], for example) in the first row and use the Fill Down (Ctrl-D) function in Excel to copy the code to all cells in the column in one fell swoop. Save the Excel file as tab-delimited text, open it in Word, and do global search and replace to put ^p before each open bracket to obtain the vertical format.

	A	B	C	D	E
1	[ENG]	submit	[POR]	apresentar	[DOM]
2	[ENG]	present	[POR]	apresentar	
3	[ENG]	offer; feature	[POR]	apresentar	
4					

Figure 7 – Converting into the vertical format.

7. WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF USING A DATABASE?

Databases are less portable than word or text files, and may require the purchase of more than one license—one for each computer. They are unlikely to be accessible from your mobile phone unless you purposely purchase one that is compatible. Interpreters who wish to access their terminology lists on the move should first consider plain text files. Depending on the database vendor, conversion to other formats may be easy, or it may be very difficult. Do some tests. If you can export to an Excel or tab-delimited file (and the export works well), then you will be able to import your data into almost any other tool. Also check how easy it is to import terminology, and which formats are allowed.

7.1 The Terminology Database in your Translation Environment Tool

Most TEnTs make it easy for you to add terms to your terminology database as you work. Some let you select the source word, then the target word, then type a control sequence to add the pair to the database. This is faster than typing the words again, if they are long words.

However, how easy is it to add terms to your terminology database when you are *not* working in the TEnT? I receive many legal and medical documents as scanned pdfs, and I then work directly in Word. Most TEnTs make it hard or impossible to add terms to the database when a project is not open.

Similarly, how easy is it to search your terminology database when you are not working in the tool? It is often hard to impossible. I keep my terminology in (vertical format) Word files and import new terms into my various TEnTs now and then — so I can have my cake and eat it too.

What percentage of your work is done in a TEnT? This is the crucial question. If you do all or most of your work in a single TEnT, that is probably the best option. However, read section 8 for information on how to search others' glossaries.

7.2 Stand-Alone Terminology Databases

© Naomi J. Sutcliffe de Moraes

Cite as: MORAES, N. J. Sutcliffe de. *Quick and Dirty: Terminology Management for the Impatient*. In "ATA 51st Annual Conference Proceedings," American Translators Association, Alexandria, VA: 2010.

For stand-alone databases, either those specifically designed for terminology or generic databases like Microsoft Access, ease of data entry depends on the tool. Likewise, ease of searching depends on the tool, to what extent it allows searching with wildcards, and if it understands inflections in your languages. These are both great advantages in Archivarius.

8. SEARCHING OTHERS' GLOSSARIES

Even if you do not use word or text files for storing the terminology you yourself research, indexing lets you leverage glossaries you find on the Internet or you receive from clients or colleagues, without adding them permanently to your own terminology records.

These come in many different formats, which may take some time to convert into whatever your favorite format is. Try indexing them. Even if the format after indexing is almost impossible to decipher, at least a search can tell you that term is in that file, and you might want to open it. Many years ago I found an interesting Portuguese/English glossary on forestry terms, published by some US government agency. I stored it away and forgot about it. More than five years later, working on a report on the Amazon forest, I discovered, when searching my glossaries, that many terms were *only* in that file. The format after indexing was horrible because it had four columns per page, but the glossary was worth the effort so I converted it using ABBYY Fine Reader. I am always finding useful glossaries in my index. Just this week I had to look up a vaccine. The source text had only an acronym. A few years ago I downloaded a monolingual glossary from Fleury, a diagnostic exam center in Brazil, with their preferred names for exams and vaccines and many alternative names commonly used by doctors. A very reliable source, and the acronym was there.

8.1 Tips

You need to create code names for your glossary files, especially if you did not create them, so you will not waste time looking in inappropriate files. I use the following format:

```
gl_ME_UNK_motorcycles_US_BR
```

Where gl stands for a bilingual glossary. I use dx for a monolingual glossary.

ME is my code for the area (mechanical engineering), as mentioned in section four. The source identifier can be UNK (unknown), telling me that I may not be able to rely on this information, or the name of a company or government body. A specific subject for a ME file could be motorcycles, or screws. I also tack on the dialects (US English and BR Portuguese in the example) when I know what they are for that file. Whatever you choose, decide on a filename format and stick with it. As you can see in figures 3 and 4, knowing which files contain which kind of information is very helpful when you have a list of 30 files, like I often do.

I have also begun to scan some paper dictionaries, especially those that are in the wrong direction — the direction I do not translate in. When a dictionary or glossary has more than one column of terms on a page, scan it using ABBYY Fine Reader, selecting each column separately (as a table) as shown in Figure 8, then saving it as formatted text or an excel file. In ABBYY, in the document language selection menu on the left, select both the source and target languages

under More Languages. I have my local bookbinder unbind the book, then use my printer's sheet feeder to quickly scan pages directly into ABBYY.

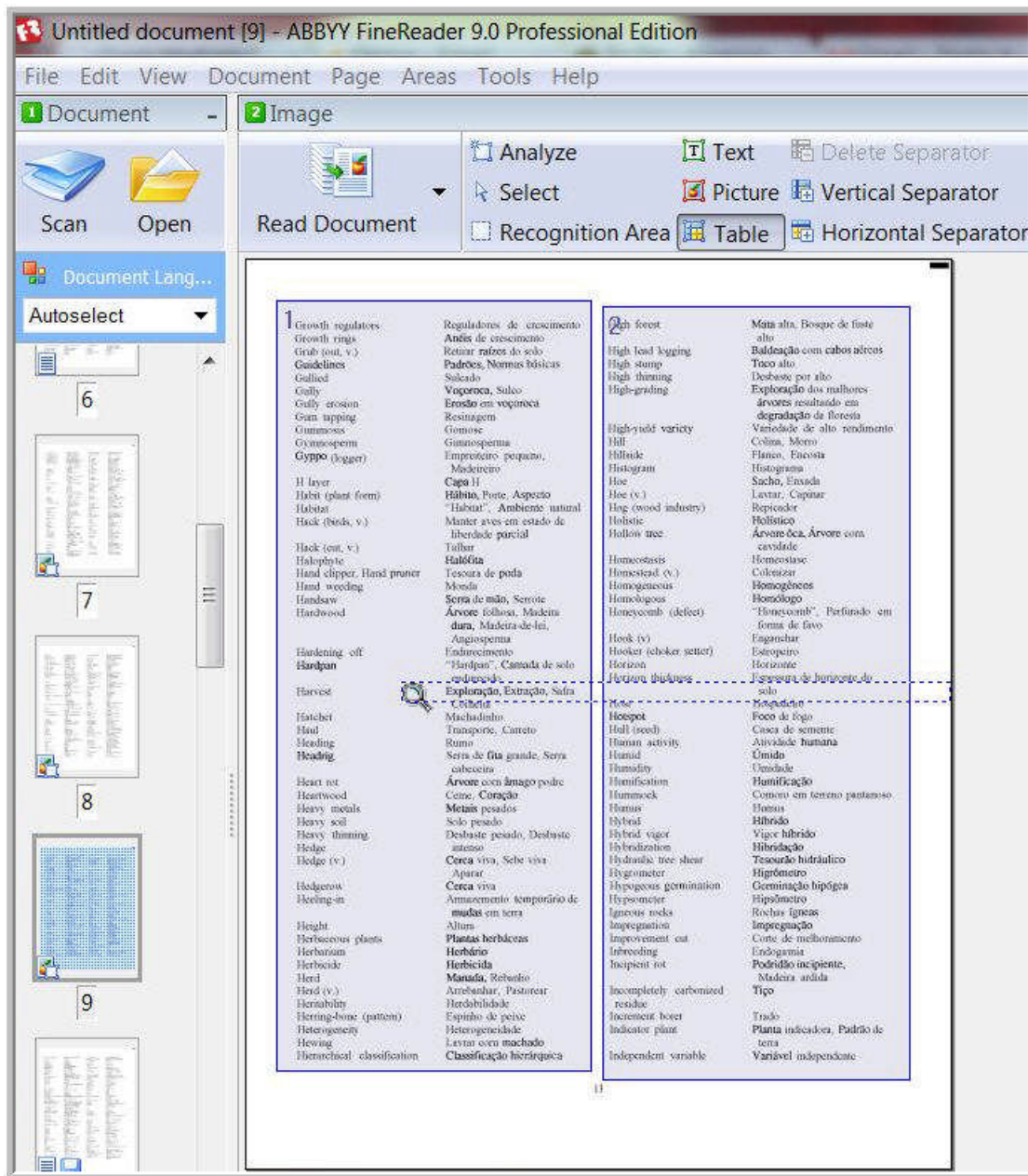


Figure 8 – Scanning two-column dictionaries with ABBYY Fine Reader

Depending on their format and your indexing software, you can also index your translation memories and bitexts. The result after indexing may or may not be useful. Do some tests.

What I have found very useful is indexing of reference materials. When translating an academic article, I look for articles in the target language on the same subject and index them. They become a ready reference when searching for terms.

© Naomi J. Sutcliffe de Moraes

Cite as: MORAES, N. J. Sutcliffe de. *Quick and Dirty: Terminology Management for the Impatient*. In "ATA 51st Annual Conference Proceedings," American Translators Association, Alexandria, VA: 2010.

9. CONCLUSIONS

I hope the discussion above has shown you how easy it is to store terminology and access it when translating. Don't reinvent the wheel with each translation. When you have found a great solution, note it down! I think many would agree that terminological knowledge is what sets the experienced translators apart from those just entering the field. Remember this:

- When you finish a job, instead of just money, you could have a list of the new terminology you discovered for future jobs. I have over 20,000 terms from 10 years of work. I rarely have to look anything up in a paper dictionary.
- Every time you stop translating to research a term for the second or nth time, you are wasting time you could be using to make more money, or spend with your loved ones, or even just by yourself reading a book.

However, like a new exercise regime, it is better to start out slow, then work up to more sophisticated tools in the future. If you choose a method that is too complicated for you, you will not stick with it. If you follow the tips above, you will start with a format that is upwardly compatible and will be able to convert your data in the future. Good luck!

10. APPENDIX

AutoHotKey script for vertical terminology input template

```
;
; AutoHotkey Version: 1.x
; Language:          English
; Platform:          Windows 7
; Author:            Naomi de Moraes
;
; Script Function:
;   Creates a template for recording terminology wherever the cursor is.
; type alt+=; (alt key then semicolon) and the template will appear
; edit with WordPad

#NoEnv      ; Recommended for performance and compatibility with future
AutoHotkey releases.
SendMode Input ; Recommended for new scripts due to its superior speed and
reliability.
SetWorkingDir %A_ScriptDir% ; Ensures a consistent starting directory.

#Singleinstance force
!;::
send
{Enter}[ENG]{Tab}{Enter}[POR]{Tab}{Enter}[DOM]{Tab}{Enter}[SOU]{Tab}{Enter}[R
EM]{Tab}{Enter}[END]{Enter}{Enter}
```

v. 2

© Naomi J. Sutcliffe de Moraes

Cite as: MORAES, N. J. Sutcliffe de. *Quick and Dirty: Terminology Management for the Impatient.* In "ATA 51st Annual Conference Proceedings," American Translators Association, Alexandria, VA: 2010.