

Dictionary Conversion from MS Excel to Abbyy Lingvo

This document describes a procedure to convert dictionaries from MS Excel into the Abbyy Lingvo format. The procedure can be used to convert any electronic dictionaries that can be copied to MS Excel or saved as TXT file with tab delimiters. The procedure uses a program¹ written by me in Python. The program sorts records alphabetically, fixes some formatting errors, creates dictionary entries, adds DSL (Lingvo's language for dictionaries) markup and produces a DSL file. The required conversion of this file to Unicode is not implemented yet and has to be done manually. My main specialty is technical translation (engineering, construction, oil and gas, etc.), experience in programming is limited, the job could be done better but everything works as designed.

For more information on my choice of formats and approach to glossaries² see Underlying Principles at the end of this manual. Below are instructions on conversion using English-Russian glossary as an example.

Preparation

The glossary you want to convert should be an Excel spreadsheet with different sheets for different language combinations: English-Russian, Russian-English and (if you have monolingual glossaries that define but do not translate) English-English and Russian-Russian.

A bilingual glossary should have exactly six columns:

1. Term.
2. Abbreviation or alternative form.
3. Translation.
4. Abbreviation or alternative form of the term in the other language.
5. Definition, explanation, examples of use.
6. Source, domain.

Example:

occupational safety and health management system	OSH management system	система управления охраной труда	СУОТ	набор взаимосвязанных или взаимодействующих между собой элементов, устанавливающих политику и цели по охране труда и достижение этих целей — A set of interrelated or interacting elements to establish OSH policy and objectives, and to achieve those objectives	ILO-OSH 2001 Guidelines on management systems for occupational health and safety // МОТ-СУОТ 2001 Руководство по системам управления охраной труда
ocean-bottom cable		донная (сейсмическая) коса			Морская сейсморазведка
ocean-bottom node	OBN	донный сейсмоприемник		Используются при сейсморазведке на море вместо донных сейсмических кос.	Морская сейсморазведка

¹ Distributed as free software.

² Glossary is a dictionary of specialized terms. There is no difference between glossaries and dictionaries from the software viewpoint.

You do not have to fill all cells, i.e. any cell in any row, except the first cell, can be empty. The first cell cannot be empty. A whole column (except the first one) can be empty but there must be at least one filled cell in the last column so that the program sees a table with six columns.

A monolingual glossary should have exactly five columns:

1. Term.
2. Abbreviation or alternative form.
3. Definition, explanation
4. Examples of use, further explanation.
5. Source, domain.

Example:

Daily Contract Quantity	DCQ	The amount of gas which a Buyer nominally undertakes to purchase and a Seller undertakes to deliver in a defined 24 hour period. Although featuring in many contracts, in practice this expression is of little meaning in itself. It may serve as a means of expressing the Annual Contract Quantity if the latter is expressed as a number of days multiplied by the DCQ. It may also define the rate at which the Seller must be able to supply gas.	See also: "Daily Delivery Rate".	Natural gas and LNG
Daily Delivery Rate		The Daily Delivery Rate (DDR) is the rate at which the Seller's facilities must be capable of delivering gas, expressed as a volume of gas per day, or as a multiple of the Daily Contract Quantity.	Also known as the Maximum Daily Quantity.	Natural gas and LNG
direct acting pressure relief valve		A pressure-relief valve that allows line pressure to act on one side of the closure element, while a predetermined spring applies a mechanical load to the other side. When the line pressure reaches its maximum limit, the line pressure overcomes the spring load and the valve opens until the line pressure falls below the preset level.	According to working principle, the safety valve or relief valve can be divided into direct-acting safety valve or relief valve and non-direct-acting safety valve or relief valve. Non-direct-acting safety valve or relief valve is divided into pilot operated safety valve or relief valve and safety valve or relief valve with auxiliary power device.	

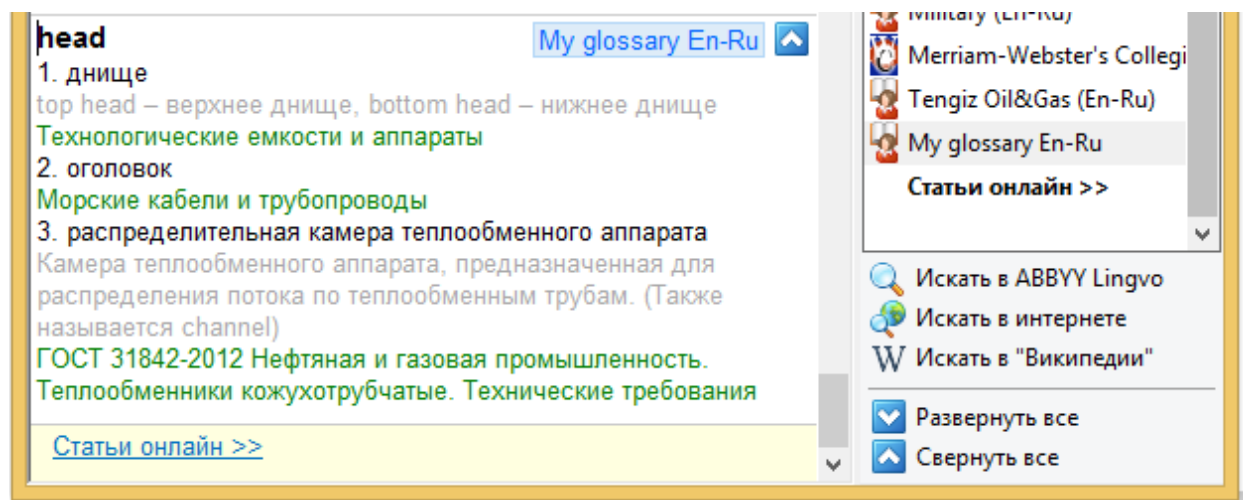
There should be **no line break or paragraph symbols in the spreadsheet as this will result in conversion errors!**

The file also **should not contain any square brackets []**, as they are used for Lingvo dictionary markup.

If you want to have different translations or definitions for the same term, you need to create several rows with identical first cells. Example:

head	1. днище	top head - верхнее днище, bottom head - нижнее днище	Технологические емкости и аппараты
head	2. оголовок		Морские кабели и трубопроводы
head	3. распределительная камера теплообменного аппарата	Камера теплообменного аппарата, предназначенная для распределения потока по теплообменным трубам. (Также называется channel)	ГОСТ 31842-2012 Нефтяная и газовая промышленность. Теплообменники кожухотрубчатые. Технические требования

Eventually, this will give you a dictionary entry with three translations for the same headword:



Translations should be numbered manually. You do not have to do it, though, but can simply add rows to the end of the spreadsheet without worrying about other rows with the same headword. The program will automatically process these rows and produce a glossary entry with different translations for the headword, although the translations will not be numbered in this case.

After that, each saved as tab delimited text.

Conversion

1. Start the program which is written in Python. To this end, you have to install Python IDE which can be downloaded from python.org. Start Python IDE, open the program file and enter data it request. The program will produce a DSL file.
2. Open the file in Wordpad and save it in Unicode (make sure file extension is dsl, not txt).
3. Convert it into LSD (Lingvo dictionary) using DSL Compiler which is included in the software package that you get when you buy Lingvo.

Underlying principles

Choice of software

I maintain my glossary in Excel as it is easy to edit and add to in this format. From time to time, I convert it into Lingvo and add to this dictionary as a user glossary.

Why not a CAT term base? Yes, computer-aided translation programs, such as Trados or Memsource, aside from translation memory, also have term bases. But I do not use CAT tools for all my translations. I also do interpreting. And I have to look up words during interpretation, though not often but nonetheless it helped me in vari-

ous settings from a drilling pad in the Arctic part of Russia to conference interpreting booth.

Lingvo provides quick full-text search (not just headwords, but works in other parts of dictionary entries). Drawback: you its versions for mobile devices do not work with user dictionaries, so on my smartphone, I use GoldenDict instead which accepts user dictionaries in the Lingvo format (lsd).

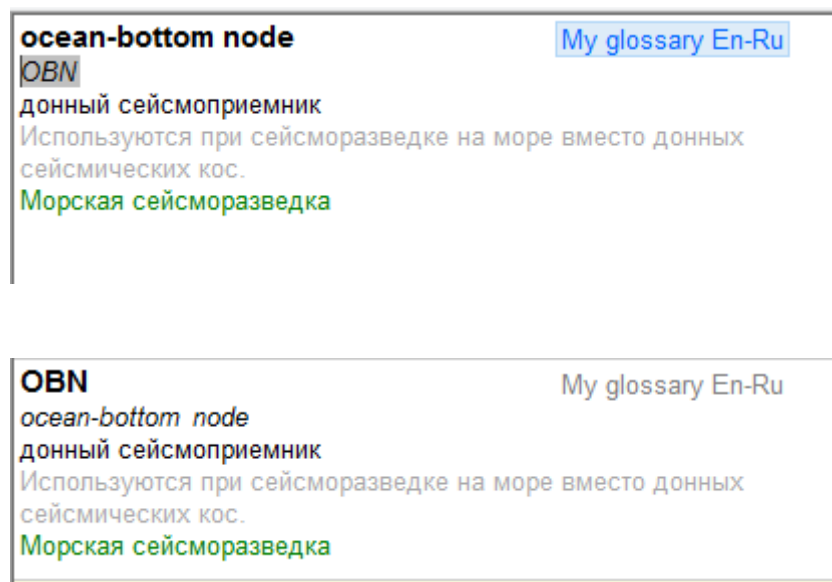
The reason I chose Python is that I am learning the language just now ☺

Glossary entry

Terms can have different meaning and translate differently depending on the domain, client and context. To produce high quality translations, one has to understand the meaning of terms, so my glossaries include not only terms and translations but also definitions, references, domains, etc.

I also maintain monolingual glossaries which only hold definitions but not translations. Standards such as ISO have lots of defined terms which proved to be very useful in my work.

If an abbreviation or alternative form was provided (see the example with “ocean bottom node – OBN” above), there will be two entries in the final user dictionary: one with the full term as the headword and the other with the abbreviation as the headword and the full term as an alternative form (see below). This will simplify search especially on mobile devices that can only search by headword.



Denis V. Kazakov. English-Russian translator/interpreter

<http://denis-kazakov.com/en.html>