Word to Tex

# 1- Introduction

Word to Tex is a Python based script which converts most information from docx files including text and header information to create a tex file. It recognizes **bold**, *italic*, underline, headings, images (if they are jpg, png, img formats), some classes of special characters (if the language is assigned), footnotes and captions. If Citavi was used, it recognizes it and adds the links. A tex template is prepended and appended to provide overall formatting.

The resulting tex file can be rendered with any tech environment (tested on tex live 🡪 texmaker on windows)

This documentation was made for GERMAN Word.

# 2- New Features in Series 1.0

Well…it works! 🡪 there are no Versions yet.

# 3 - Installing Word to Tex

Word to Tex does not need to be installed. Unpack all files into a directory of your choice. The files contained in the package should include

* word2tex.exe – the executable
* huplc\_pre.tex – the default tex template prepended to the resulting tex file
* huplc\_post.tex – the default tex template appended to the resulting tex file

Additionally, you need a Tex environment to render the resulting tex file.

# 4 – Installing Tex Life (2018)

Install Tex Life (2018). Use all default options.

Site: <https://www.tug.org/texlive/>

Download: <http://mirror.ctan.org/systems/texlive/tlnet/install-tl-windows.exe>

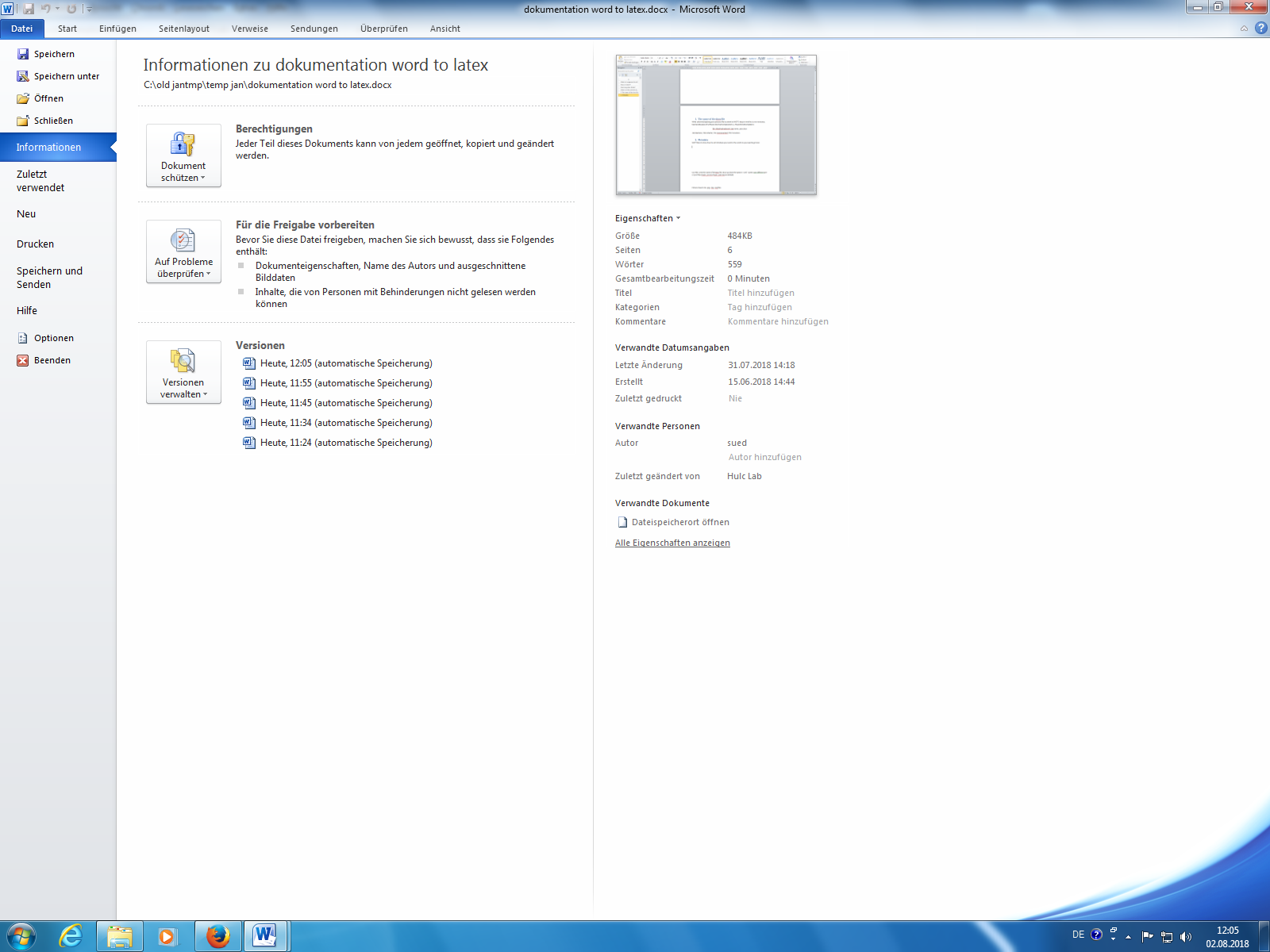
(Last checked 09.08.2018)

We used Texworks for testing (part of tex life 2018)

# 5- Before Word to Tex can be used

## 5.1 Prepare the docx file

You need to add Metadata

The needed information is: Title, comments, categories and author.

Under comments you add information about you institute, history and language.

The used format is:

--institute= XXXX --history= date of submission --language=XXXXX

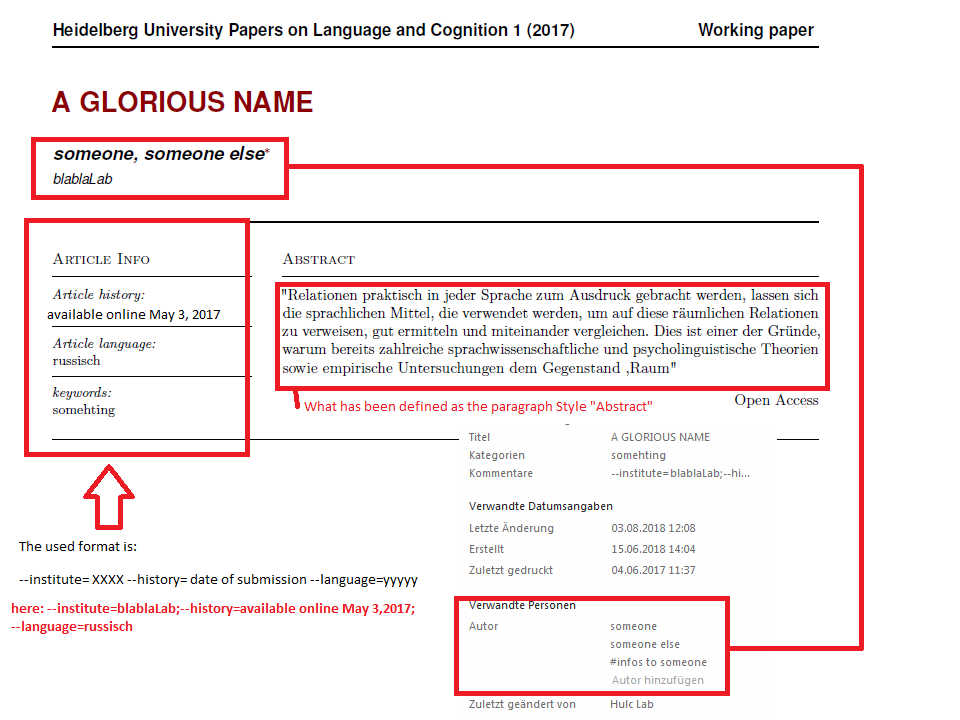
--articletype=XXXX --pub=XXXX

Articletype = what kind of article (default = Working paper)

pub = Publication name

Here an example how the metadata is used:

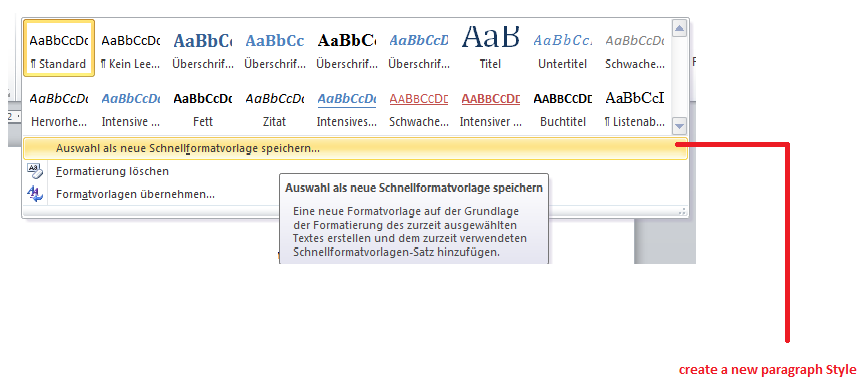




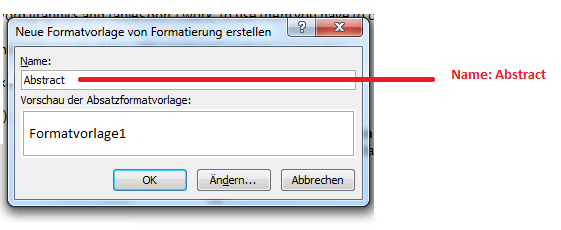
## 5.2 Abstract

To define the abstract you need to create a new paragraph Style and name it “Abstract”. This paragraph will be placed into the abstract placeholder.

1)

2)

3)



4) Assign it to your abstract

## 5.3 Pictures and Graphics

Pictures have to be .jpg, .img, .png files

Word graphics and tables don’t work, to use them you have to change them to pictures.

The same can be done with constructs made by using tabs:

Example:

1. *„Stojím před domem.“*

st-ojím *před*  d-**omem**

steh-1.Ps.Sg.Präs. vor Haus-Mask.Sg.**Instr.**

„Ich stehe vor dem Haus.“

*vs.*

*„Běž před dům!“*

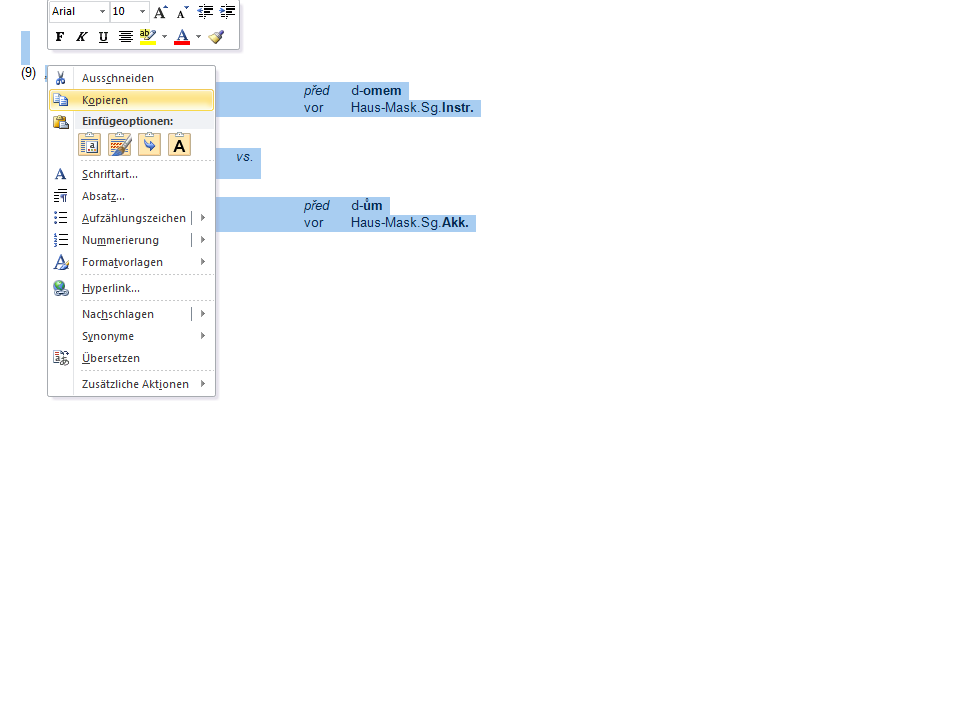
bě-ž *před* d-**ům**

lauf-2.Ps.Sg.Präs.Imperativ. vor Haus-Mask.Sg.**Akk.**

„Lauf vor das Haus!“

STEP 1:

Word 2010 🡪 r click copy (or select and press control c)

**

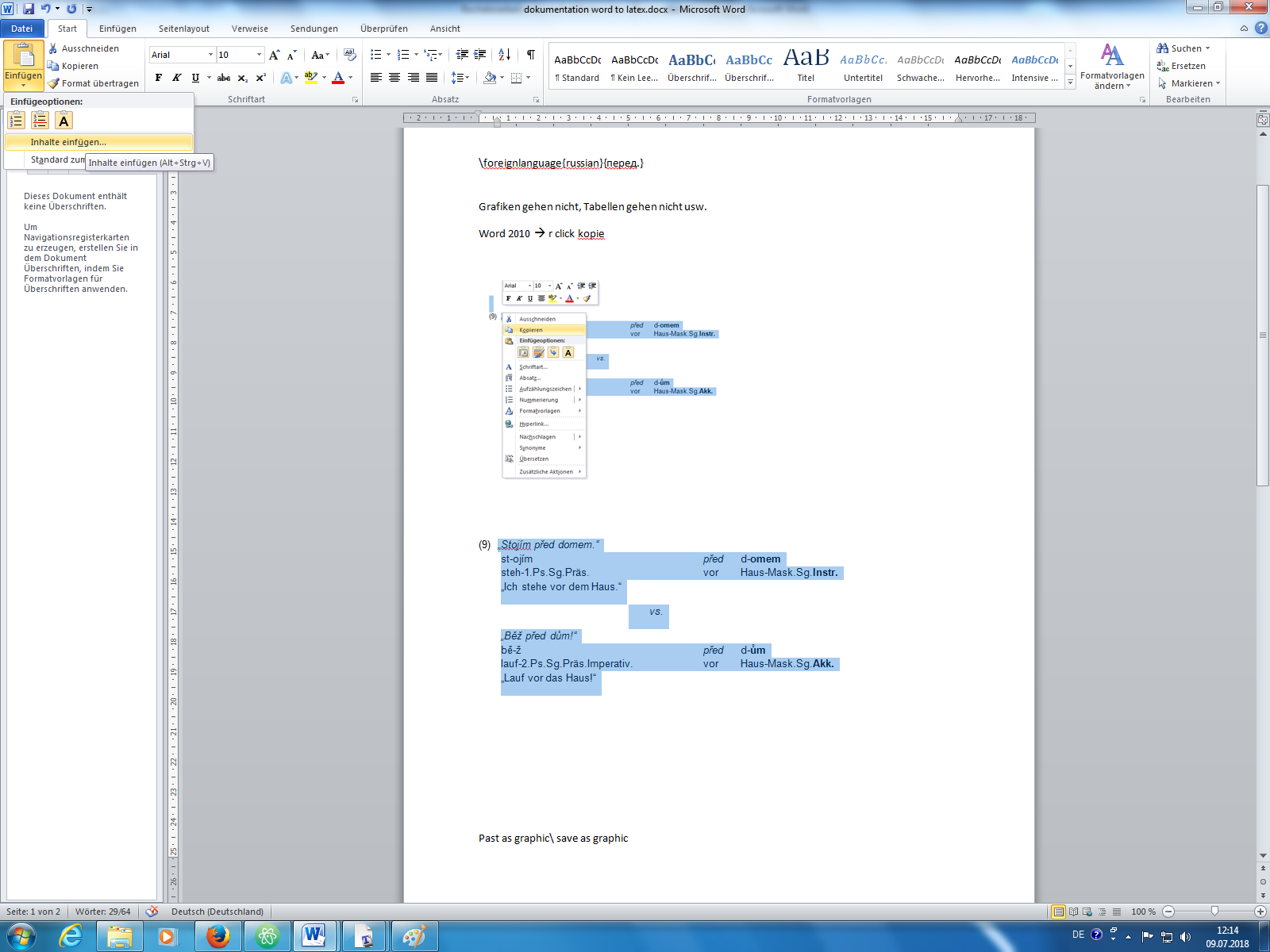
Step 2:

Start (or Home in eng.)

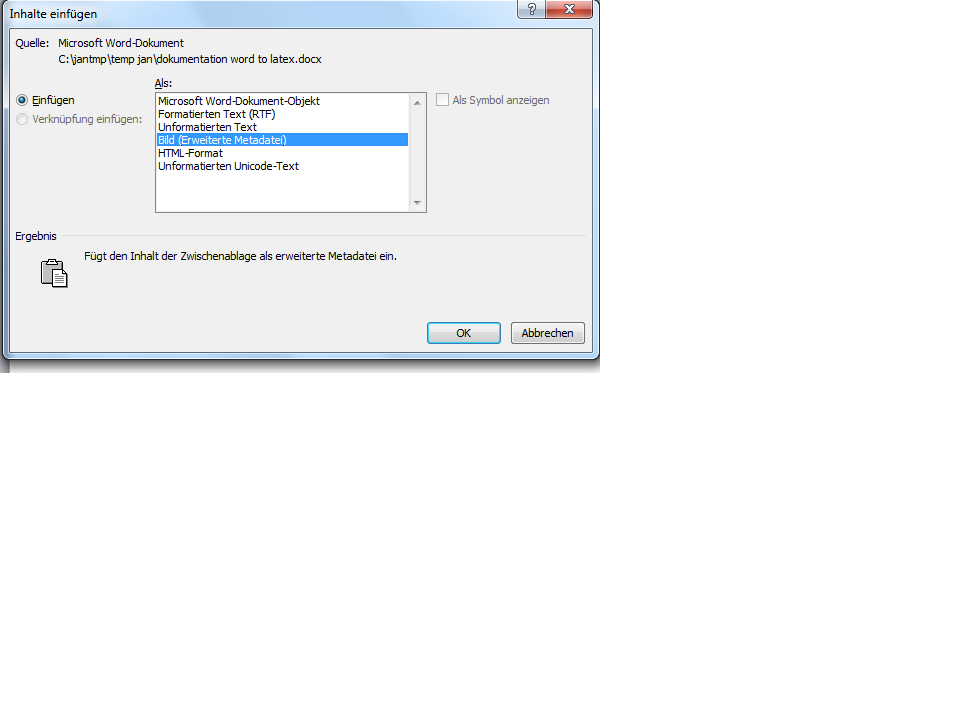
Einfügen\Paste 🡪 underneath are more options

Ger: Inhalte einfügen,

Eng: special past options (or something similar)

**

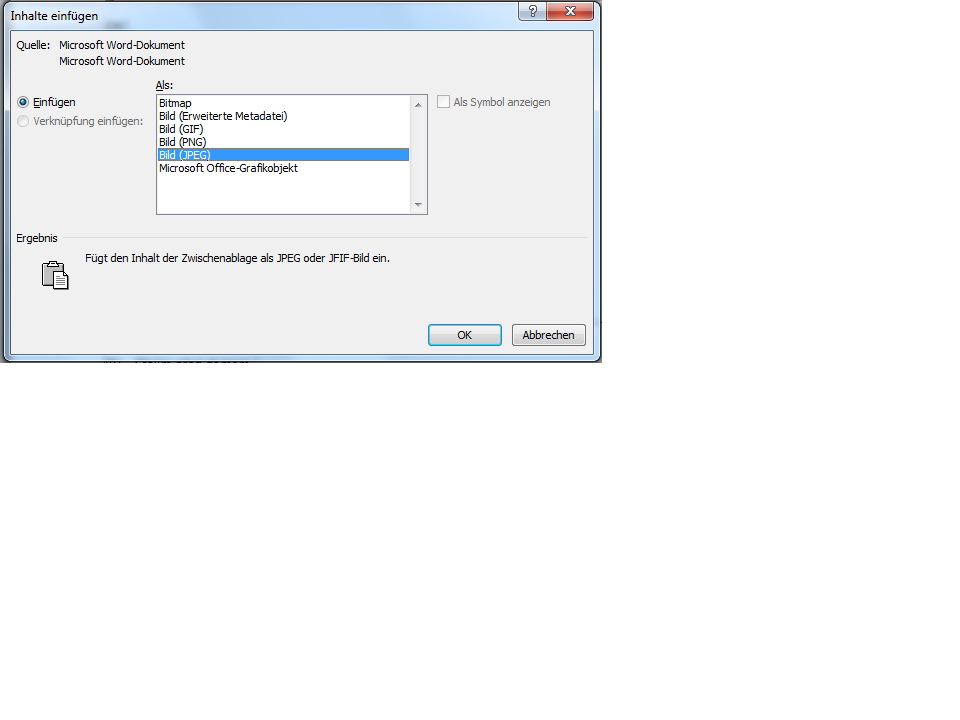
Step3: Chose Bild (Erweiterte Metadatei)\ Chose Picture\graphics



OK!

IT’S A PICTURE!... but it’s a .emf picture…. LATEX can’t use this format… so:

Step 1 + 2 again, but this time the window that appears reads:



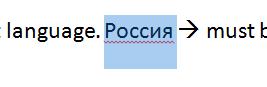
Chose JPEG!

Now you have the picture twice! 🡪 Delete the bottom one… it’s the “old” .emf file!

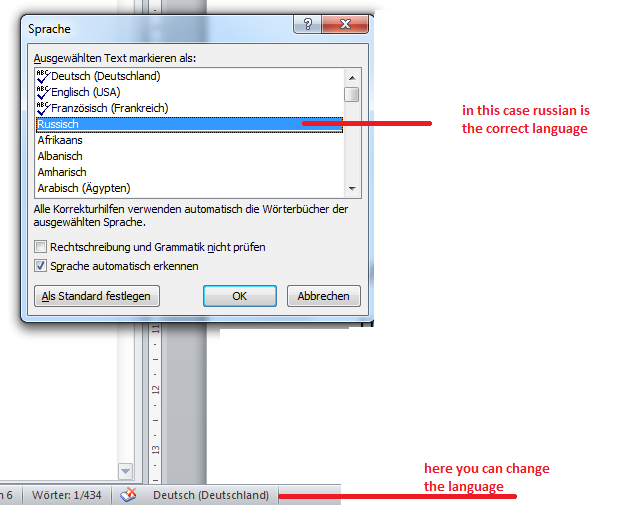
## 5.4 Special characters

Special characters must be assigned their correct language. Россия 🡪 must be assigned Russian.

1)



2)



## 5.5 Bibliography + Citavi

Full explanation:

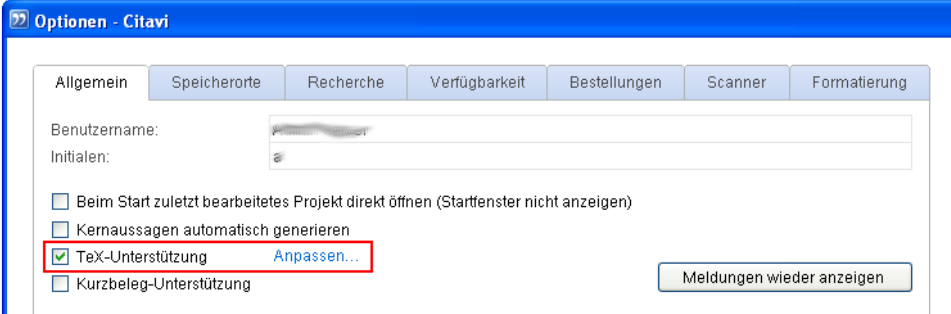
http://www.suedraum.de/latex/stammtisch/degenkolb\_latex\_biblatex\_folien-final.pdf

1) The bibliography is created automatically by using Citavi and biber

2) If a bibliography is already present, it has to be deleted.

3) Use Citavi

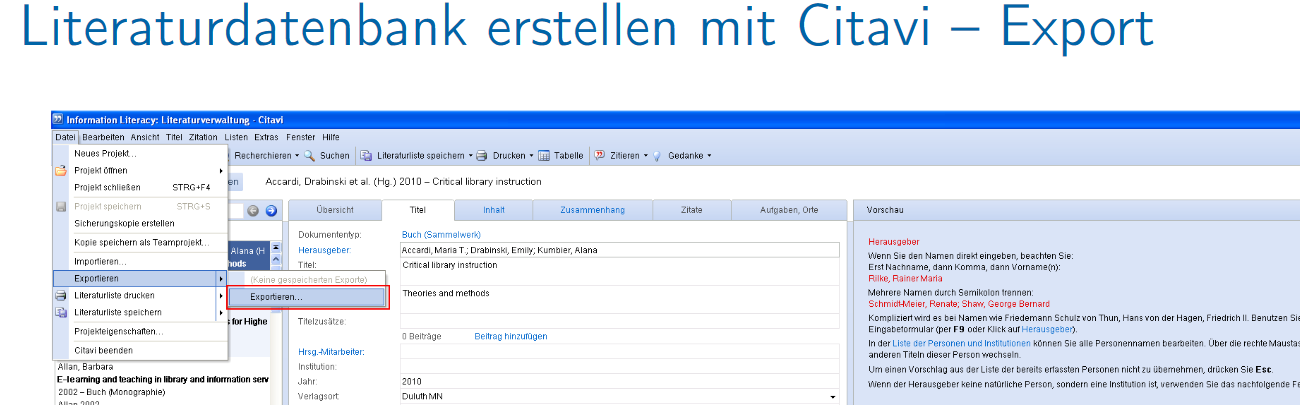
4) You have to activate bibtex mode in citavi!



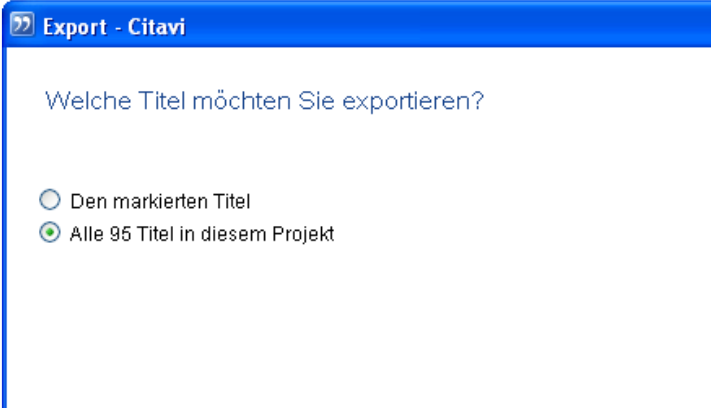
5) It’s possible that you need to assign the in text references again

6) Export the library as a bib.file

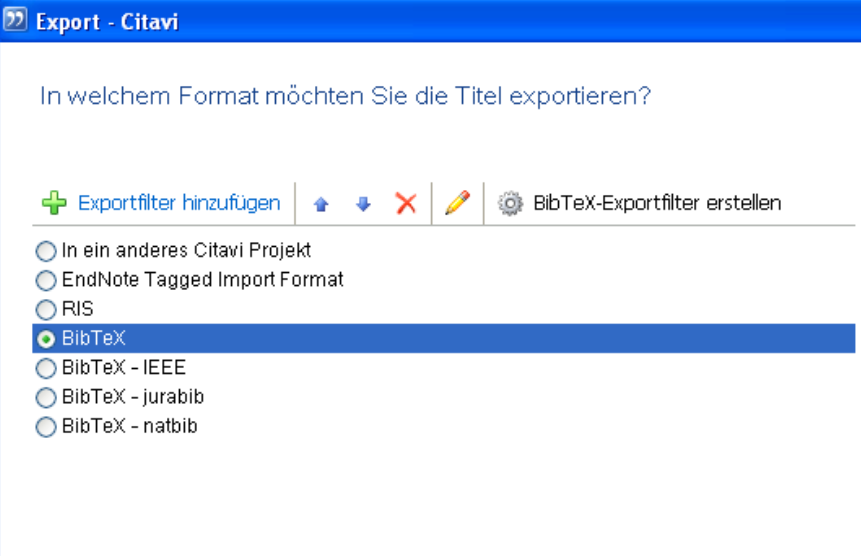
a)



b) Export all titles (in this example its 95 titles)



c) Choose BibteX format



Create the bib file, use it later.

## 5.6 Special cases

1) Delete things like a line between chapters made out of “–“or “\_”.

Like this:

It will cause trouble. (Like lines going over the whole page, even if there is other text or images)

# 6 – Usage of word2tex

Use the command line to run Word2Tex.exe and afterwards the docx file name.

All files must be in the same folder (so word2tex.exe, huplc\_pre.tex, huplc\_post.tex, the .bib file if you have one and your docx file)

Options must be written before your docx filename.

Examples:

C:\yourfolder\Word2tex.exe --bibfile test.bib test.docx

C:\myhovercraftisfullofeels\Word2tex.exe test.docx – this will use no bib file and the default huplc\_pre.tex and huplc\_post.tex

C:\yourfolder\Word2tex.exe --pre mytemplate.tex test.docx - no bib file, mytemplate.tex as pre file

This will create: a .tex file with the name of your docx files so test.docx 🡪 test.tex

## 6.1 If you have a bib file

If you have a bib file, you need to delete any existing Bibliography, because a new one will be created. Than use biber, the tex bibliography tool, to create a bibliography and link the intext citations. After this is done, use the normal pdflatex to create your pdf file.