Umfang und Aufbau des Korpus

- Konkretere Informationen zur Korpusaufbau und Korpusgröße in Bezug auf die Originalquellen
- v 0.1
- v 0.2

Konkretere Informationen zur Korpusaufbau und Korpusgröße in Bezug auf die Originalquellen

In der folgenden Tabelle werden die bearbeiteten Textausschnitte in Relation zu einem Schlagwortkorpus, das das gesamte 19. und das frühe 20. Jahrhundert (bis 1918) grob abdeckt und nachvollziehbar macht, gesetzt. Insgesamt können (am 17. 05. 2021) mit der Suchabfrage **"utra??ist*" im ANNO bis inklusive 1918 6.338 Treffer** (= Ausgaben von Zeitungen und Zeitschriften, in denen die Suchabfrage mindestens ein Mal vorkommt) gefunden werden, von denen **6.179** auf Zeitungen und 159 auf Zeitschriften entfallen. Letztere wurden bei den Überlegungen zu einem möglichen Gesamtkorpus außer Acht gelassen.

Periode 1800–1829

Publikationsort	Link zu ANNO	Treffer	davon im Korpus
Brno	Link	**15**	**13** (2 Duplikate)
Wien	Link	**11**	**11**
L'viv	Link	1	**0** (weil Duplikat)

Periode 1850–1859

Publikationsort	Link zu ANNO	Treffer	davon im Korpus
Wien	Link	**29**	**8**

Olomouc	Link	3	**3**
Salzburg	Link	3	**3**
Brno	Link	1	**1**
Opava	Link	1	**1**
Litoměřice	Link	1	**0**

Periode 1867–1868 (stellvertretend für die 1860er)

Jahr	Publikationsort	Link zu ANNO	Treffer	davon im Korpus
1867	Wien	Link	**13**	**11**
1867	Plzeň	Link	4	**0**
1867	Graz	Link	1	**0**
1867	Klagenfurt	Link	1	**0**
1867	Znojmo	Link	1	**0**
1867	Olomouc	Link	4	**0**
1867	Praha	Link	2	**0**
1868	Wien	Link	**22**	**8**
1868	Znojmo	Link	7	**0**
1868	Praha	Link	6	**0**
1868	Graz	Link	1	**0**
1868	Plzeň	Link	1	**0**

Jahr 1890

Publikationsort	Zeitung	Link zu ANNO	Treffer	davon im Korpus
Wien	Neue Freie Presse	Link	**11**	**8**
Wien	Die Presse	Link	8	**8**
Wien	Das Vaterland	Link	**12**	**10**
Wien	Wiener Montags-Journal	Link	1	**0**
Wien	Wiener Zeitung	Link	6	**6**
Wien	Deutsches Volksblatt	Link	4	**4**
Praha	alle	Link	**13**	**0**
Maribor	alle	Link	9	**6**
Salzburg	alle	Link	4	**4**
Litoměřice	alle	Link	8	**0**
Linz	alle	Link	7	**0**
Znojmo	alle	Link	3	**0**

Jahr 1910

Publikationsort	Zeitung	Link zu ANNO	Treffer	davon im Korpus
Wien	Neue Freie Presse	Link	**13**	**10**
Wien	Wiener Zeitung	Link	**13**	**10**
Wien	Neues Wiener Tagblatt	Link	6	**6**
Wien	Zollämter- und Finanzwachzeitung	Link	4	**4**

Wien	Deutsches Volksblatt	Link	5	**5**
Wien	Arbeiterzeitung	Link	3	**3**
Praha	alle	Link	**30**	**0**
Maribor	alle	Link	**15**	**6**
Klagenfurt	alle	Link	**10**	**9**
Graz	alle	Link	**24**	**10**
Czernowitz	alle	Link	**13**	**0**

Mitarbeit?

Sie haben Interesse, einzelne Ausgaben zu einem der Datenpakete zu ergänzen? Gerne sende ich Ihnen eine genaue Liste der noch fehlenden Ausgaben zu! Melden Sie sich bei: agnes.kim@univie.ac.at

v 0.1

Veröffentlichung

Datum: 12.05.2021

Link: https://dioecloud.trans.univie.ac.at/index.php/s/36iS7xZ4f8dD6fK

Größe des Korpus

Textausschnitte: 156

Tokens: 31,176

Wörter: 26,806

Überblick über den Korpusaufbau

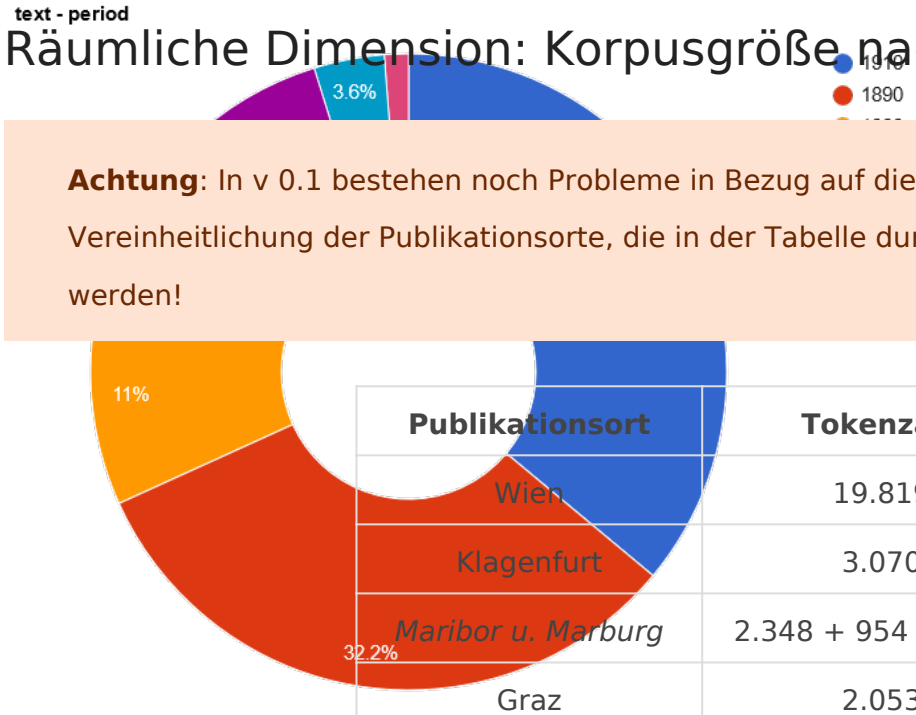
Die Grafiken beziehen sich auf die Gesamtzahl der Token. Sie wurden mit SketchEngine erstellt.

Zeitliche Dimension: Korpusgröße nach Jahrzehnten

Jahrzehnt	Tokenzahl
1800	382
1820	1.115
1850	2.343
1860	3.419
1890	10.024

1900	2.650
1910	11.241

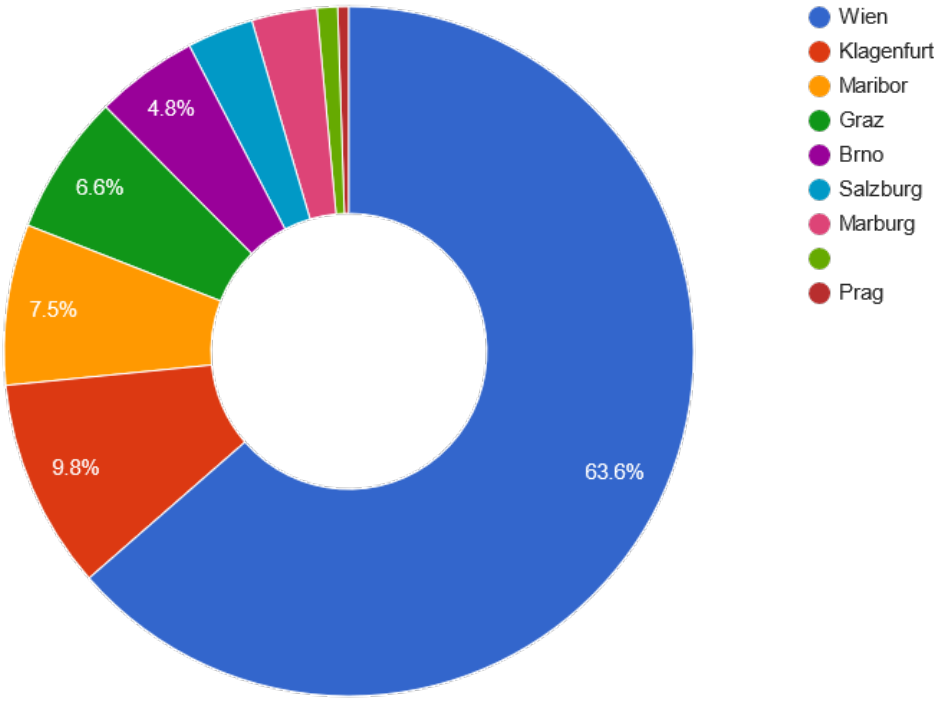
Räumliche Dimension: Korpusgröße nach Publikationsorten



Achtung: In v 0.1 bestehen noch Probleme in Bezug auf die Zuweisung bzw. Vereinheitlichung der Publikationsorte, die in der Tabelle durch Kursivsatz hervorgehoben werden!

Publikationsort	**Tokenzahl**
Wien	19.819
Klagenfurt	3.070
Maribor u. Marburg	2.348 + 954 = 3.202
Graz	2.053
Brno	1.497
Salzburg	971
nicht zugewiesen	299
Prag	163

text - place


4.8%
6.6%
7.5%
9.8%
63.6%
Wien
Klagenfurt
Maribor
Graz
Brno
Salzburg
Marburg
Prag

v 0.2

Veröffentlichung

Datum: 06.06.2021

Link: https://dioecloud.trans.univie.ac.at/index.php/s/P7p2jMcoFMgWFbx

Änderungsnotizen

- Vereinheitlichung der Periodenzuordnung (`period`)
- Vereinheitlichung der Publikationsorte (`place`)
- Vereinheitlichung der Annotation der orthographischen und morphologischen Normalisierungen (`edit type`)

Größe des Korpus

Textausschnitte: 156

Tokens: 31,176

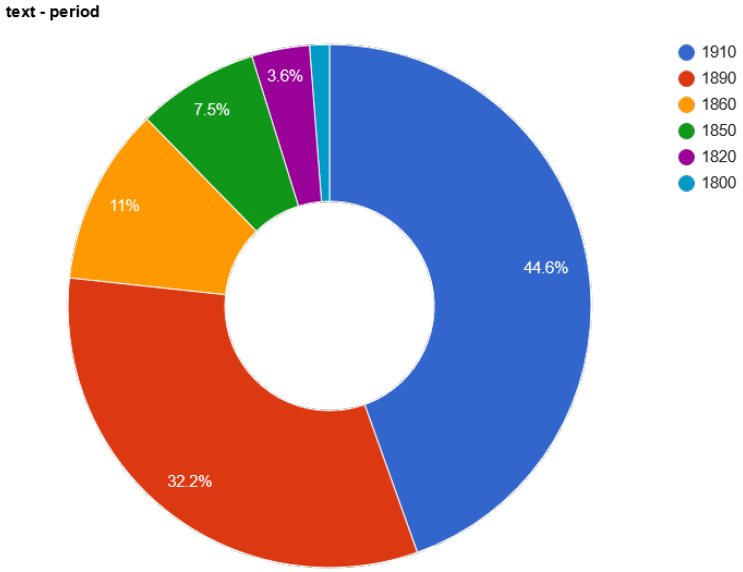
Wörter: 26,806

Überblick über den Korpusaufbau

Die Grafiken beziehen sich auf die Gesamtzahl der Token. Sie wurden mit SketchEngine erstellt.

Zeitliche Dimension: Korpusgröße nach Jahrzehnten

Jahrzehnt	Tokenzahl
1800	382
1820	1.115
1850	2.343
1860	3.419
1890	10.024
1910	13.888



Räumliche Dimension: Korpusgröße nach Publikationsorten

Achtung: In v 0.1 bestehen noch Probleme in Bezug auf die Zuweisung bzw. Vereinheitlichung der Publikationsorte, die in der Tabelle durch Kursivsatz hervorgehoben werden!

Publikationsort	Tokenzahl
Wien	20.119
Maribor	3.302
Klagenfurt	3.070
Graz	2.053
Brno	1.497
Salzburg	971
Prag	163

