

# Collocations in parallel sentences

November 5, 2020

```
[1]: import os
import math
import lzma
```

Download corpus via `wget -r -np https://pub.cl.uzh.ch/corpora/PaCoCo/Rumantsch_Grischun/`

```
[2]: PATH = 'pub.cl.uzh.ch/corpora/PaCoCo/Rumantsch_Grischun'
TOKEN_FILES = 'rumantsch-grischun.token.{}.tsv.xz'
SENTALIGN_FILES = 'rumantsch-grischun.sentence_alignment.{}.tsv.xz'
LANGUAGES = ['de', 'rm']
```

This function reads from compressed files directly:

```
[3]: def readxzfile(filename, language):
    """Reads in a compressed file in PaCoCo format (extended CoNLL-U) and
    →returns rows as lists"""
    fh = lzma.open(filename.format(language), mode='rt', encoding='utf-8')
    for line in fh:
        if line == '\n' or line.startswith('#'):
            continue
        yield line.rstrip().split('\t')
```

Read in sentence alignment units (relation sentenceID <=> sentenceAU):

```
[4]: sentAUs = {}
for language in LANGUAGES:
    for au in readxzfile(os.path.join(PATH, SENTALIGN_FILES), language):
        sentAUs.setdefault(int(au[1]), {}).setdefault(language, []).
        →append(int(au[0]))
```

Read in token lists (relation token <=> sentenceID):

```
[5]: tokens = {}
for language in LANGUAGES:
    for au in readxzfile(os.path.join(PATH, TOKEN_FILES), language):
        tokens.setdefault(int(au[1]), []).append(au[1])
```

Let's have a look at an example:

```
[6]: sentAUs[1000]
```

```
[6]: {'de': [10039484], 'rm': [160039414]}
```

```
[7]: ' '.join(tokens[10039484])
```

```
[7]: '^2 Jeder Indossant , der den Wechsel eingelöst hat , kann sein Indossament und  
die Indossamente seiner Nachmänner austreichen .'
```

```
[8]: ' '.join(tokens[160039414])
```

```
[8]: '^2 Mintga indossant che ha pajà la cambiala po stritgar ses indossament e quels  
dals indossants posteriurs .'
```

We iterate over all tokens in both languages and calculate frequencies and a cooccurrence matrix:

```
[9]: tlist = {}  
pairs = []  
freq = {}  
coo = {}  
for sentAU in sentAUs:  
    termlist = {}  
    for language in LANGUAGES:  
        tlist[language] = [t for id in sentAUs[sentAU][language] for t in  
→tokens[id]]  
        termlist[language] = set([w.lower() for w in tlist[language]])  
        for w in termlist[language]:  
            freq.setdefault(language, {}).setdefault(w, 0)  
            freq[language][w] += 1  
        # skip sentences with less than 3 tokens  
        if len(tlist['de']) < 3 or len(tlist['rm']) < 3:  
            continue  
        # cooccurrence matrix  
        for w1 in termlist['de']:  
            for w2 in termlist['rm']:  
                coo.setdefault(w1, {}).setdefault(w2, 0)  
                coo[w1][w2] += 1  
    pairs.append(tlist.copy())
```

The total number of tokens per language:

```
[10]: total = {}  
for language in LANGUAGES:  
    total[language] = sum(list(freq[language].values()))
```

```
[11]: total
```

```
[11]: {'de': 381932, 'rm': 454399}
```

The ten most frequent tokens per language (lowercased):

```
[12]: sorted(freq['de'].items(), key=lambda item: item[1], reverse=True)[:10]
```

```
[12]: [('.', 14617),
      ('die', 11219),
      ('der', 10516),
      ('', 9728),
      ('art.', 7274),
      ('und', 7198),
      ('oder', 5171),
      ('des', 4759),
      ('^1', 3973),
      ('^2', 3961)]
```

```
[13]: sorted(freq['rm'].items(), key=lambda item: item[1], reverse=True)[:10]
```

```
[13]: [('da', 14640),
      ('.', 14517),
      ('la', 11573),
      ('', 8319),
      ('il', 7737),
      ('l"', 7668),
      ('art.', 7288),
      ('en', 6157),
      ('e', 6102),
      ('las', 5640)]
```

Ranking of the cooccurrences using mutual information (MI) score:

```
[14]: ranking = {}
      for w1, w2list in coo.items():
          for w2, count in w2list.items():
              o = count
              e = freq['de'][w1] * freq['rm'][w2] / min(total.values())
              # skip any hapax legomenon
              if o == 1:
                  continue
              # skip any negative association
              if o < e:
                  continue
              # MI is calculated bitwise (base=2)
              score = math.log(o/e, 2) # MI
              ranking[(w1, w2)] = score
```

Pairs with the highest mutual information:

```
[15]: sorted(ranking.items(), key=lambda item: item[1], reverse=True)[:1000]
```

[15]: [ (('allmächtigen', 'tutpussant'), 17.542956275013225),  
 (('frieden', 'conscients'), 17.542956275013225),  
 (('bestreben', 'conscients'), 17.542956275013225),  
 (('offenheit', 'avertadad'), 17.542956275013225),  
 (('zürich', 'turitg'), 17.542956275013225),  
 (('wallis', 'vallais'), 17.542956275013225),  
 (('5a', '5a'), 17.542956275013225),  
 (('grausamer', 'crudaivel'), 17.542956275013225),  
 (('unmenschlicher', 'crudaivel'), 17.542956275013225),  
 (('gewissensfreiheit', 'conscienza'), 17.542956275013225),  
 (('gewissensfreiheit', 'cretta'), 17.542956275013225),  
 (('fernmeldetechnischen', 'telediffusiun'), 17.542956275013225),  
 (('unverfälschte', 'nunfalsifitgada'), 17.542956275013225),  
 (('bedrohungen', 'guerras'), 17.542956275013225),  
 (('61a', '61a'), 17.542956275013225),  
 (('abschlüssen', 'passadis'), 17.542956275013225),  
 (('herkömmliche', 'tradiziunalas'), 17.542956275013225),  
 (('niederschläge', 'precipitaziuns'), 17.542956275013225),  
 (('stauanlagen', 'serra'), 17.542956275013225),  
 (('stätten', 'istorics'), 17.542956275013225),  
 (('pflanzenwelt', 'flora'), 17.542956275013225),  
 (('pflanzenwelt', 'fauna'), 17.542956275013225),  
 (('fische', 'peschs'), 17.542956275013225),  
 (('a-c', 'a-c'), 17.542956275013225),  
 (('konsumentinnen', 'consumentas'), 17.542956275013225),  
 (('streitwertgrenze', 'lita'), 17.542956275013225),  
 (('steuererleichterungen', 'levgiaments'), 17.542956275013225),  
 (('chemikalien', 'chemicalias'), 17.542956275013225),  
 (('waffenzubehör', 'muniziu'), 17.542956275013225),  
 (('waffenzubehör', 'accessoris'), 17.542956275013225),  
 (('munition', 'muniziu'), 17.542956275013225),  
 (('munition', 'accessoris'), 17.542956275013225),  
 (('112b', '112b'), 17.542956275013225),  
 (('betagten-', '112c'), 17.542956275013225),  
 (('112c', '112c'), 17.542956275013225),  
 (('keimgut', 'schermentzas'), 17.542956275013225),  
 (('zellen', 'tessids'), 17.542956275013225),  
 (('geweben', 'tessids'), 17.542956275013225),  
 (('pubertät', 'pubertad'), 17.542956275013225),  
 (('messwesen', 'metrologia'), 17.542956275013225),  
 (('141a', '141a'), 17.542956275013225),  
 (('versammeln', 'radunan'), 17.542956275013225),  
 (('abgabepflichtigen', 'contribuents'), 17.542956275013225),  
 (('neutralität', 'neutralitad'), 17.542956275013225),  
 (('aktivdienst', 'mobilisescha'), 17.542956275013225),  
 (('regierungspolitik', 'guvernamentalala'), 17.542956275013225),  
 (('t', 'paisa'), 17.542956275013225),

(( 't', 'tonnas'), 17.542956275013225),  
(( 'ost-', 'vest'), 17.542956275013225),  
(( 'bahn', 'vest'), 17.542956275013225),  
(( 'eisenbahn-hochleistungsnetz', 'vest'), 17.542956275013225),  
(( 'westschweiz', 'vest'), 17.542956275013225),  
(( 'gesamtaufwendungen', 'totalas'), 17.542956275013225),  
(( '2020', '2020'), 17.542956275013225),  
(( 'uno', 'onu'), 17.542956275013225),  
(( 'gewaltenhemmung', 'pussanzas'), 17.542956275013225),  
(( 'gewaltenteilung', 'pussanzas'), 17.542956275013225),  
(( 'bündnerischen', 'grischuns'), 17.542956275013225),  
(( 'bündnerischen', 'grischunas'), 17.542956275013225),  
(( 'bezirksgerichte', 'districtualas'), 17.542956275013225),  
(( 'kreispräsidentinnen', 'substituts'), 17.542956275013225),  
(( 'kreispräsidentinnen', 'circuitals'), 17.542956275013225),  
(( 'kreispräsidentinnen', 'substitutas'), 17.542956275013225),  
(( 'aussergerichtliche', 'extragiudizialas'), 17.542956275013225),  
(( 'monopole', 'monopols'), 17.542956275013225),  
(( 'konfessionell', 'confessiunalmain'), 17.542956275013225),  
(( 'konfessionell', 'politicamain'), 17.542956275013225),  
(( 'politisch', 'confessiunalmain'), 17.542956275013225),  
(( 'politisch', 'politicamain'), 17.542956275013225),  
(( 'neutral', 'confessiunalmain'), 17.542956275013225),  
(( 'neutral', 'politicamain'), 17.542956275013225),  
(( 'evangelisch-reformierte', 'catholic-romana'), 17.542956275013225),  
(( 'evangelisch-reformierte', 'evangelic-refurmada'), 17.542956275013225),  
(( 'wandelt', 'transfurma'), 17.542956275013225),  
(( 'efta', 'aecl'), 17.542956275013225),  
(( 'beschwerdefähige', 'codex'), 17.542956275013225),  
(( 'beschwerdefähige', 'v'), 17.542956275013225),  
(( 'grenzkodex', 'codex'), 17.542956275013225),  
(( 'grenzkodex', 'v'), 17.542956275013225),  
(( 'v', 'codex'), 17.542956275013225),  
(( 'v', 'v'), 17.542956275013225),  
(( 'termingerechte', 'produceschian'), 17.542956275013225),  
(( 'hochstehende', 'produceschian'), 17.542956275013225),  
(( 'garantieren', 'produceschian'), 17.542956275013225),  
(( '42-44', '42-44'), 17.542956275013225),  
(( 'ausländerorganisationen', 'nunguvernamentalas'), 17.542956275013225),  
(( 'nichtregierungsorganisationen', 'nunguvernamentalas'), 17.542956275013225),  
(( '1954', '1954'), 17.542956275013225),  
(( '71a', '71a'), 17.542956275013225),  
(( 'besessen', 'possedi'), 17.542956275013225),  
(( '75-77', '75-77'), 17.542956275013225),  
(( '81.', '81.'), 17.542956275013225),  
(( 'konkret', 'concretamain'), 17.542956275013225),  
(( 'verletze', 'violeschia'), 17.542956275013225),

(('98a', '98a'), 17.542956275013225),  
 (('flughafenbetreiber', 'gestiunaris'), 17.542956275013225),  
 (('katalog', 'catalog'), 17.542956275013225),  
 (('109d', '109d'), 17.542956275013225),  
 (('dokumentationssystem', 'dossiers'), 17.542956275013225),  
 (('personendossier-', 'dossiers'), 17.542956275013225),  
 (('allianz-', 'allianza'), 17.542956275013225),  
 (('allianz-', 'implantats'), 17.542956275013225),  
 (('allianz-', 'artista'), 17.542956275013225),  
 (('allianz-', 'protesas'), 17.542956275013225),  
 (('kennzeichen', 'allianza'), 17.542956275013225),  
 (('kennzeichen', 'implantats'), 17.542956275013225),  
 (('kennzeichen', 'artista'), 17.542956275013225),  
 (('kennzeichen', 'protesas'), 17.542956275013225),  
 (('implantate', 'allianza'), 17.542956275013225),  
 (('implantate', 'implantats'), 17.542956275013225),  
 (('implantate', 'artista'), 17.542956275013225),  
 (('implantate', 'protesas'), 17.542956275013225),  
 (('ordens-', 'allianza'), 17.542956275013225),  
 (('ordens-', 'implantats'), 17.542956275013225),  
 (('ordens-', 'artista'), 17.542956275013225),  
 (('ordens-', 'protesas'), 17.542956275013225),  
 (('prothesen', 'allianza'), 17.542956275013225),  
 (('prothesen', 'implantats'), 17.542956275013225),  
 (('prothesen', 'artista'), 17.542956275013225),  
 (('prothesen', 'protesas'), 17.542956275013225),  
 (('111a', '111a'), 17.542956275013225),  
 (('111b', '111b'), 17.542956275013225),  
 (('b-d', 'b-d'), 17.542956275013225),  
 (('111d', '111d'), 17.542956275013225),  
 (('111f', '111f'), 17.542956275013225),  
 (('1961', '1961'), 17.542956275013225),  
 (('2^ter', '2^ter'), 17.542956275013225),  
 (('2^quater', '2^quater'), 17.542956275013225),  
 (('paralleldatensammlungen', 'parallelas'), 17.542956275013225),  
 (('summarisch', 'summaricamain'), 17.542956275013225),  
 (('ja-', '""na""'), 17.542956275013225),  
 (('ja-', '""gea""'), 17.542956275013225),  
 (('nein-stimmen', '""na""'), 17.542956275013225),  
 (('nein-stimmen', '""gea""'), 17.542956275013225),  
 (('gruppierungen', 'gruppaziuns'), 17.542956275013225),  
 (('präsidierenden', 'presidieschan'), 17.542956275013225),  
 (('wahlanleitung', 'mussavia'), 17.542956275013225),  
 (('vorgedruckte', 'prestampà'), 17.542956275013225),  
 (('quotienten', 'quozient'), 17.542956275013225),  
 (('unzulässigkeit', 'inadmissiblidad'), 17.542956275013225),  
 (('herunterlädt', 'telechargia'), 17.542956275013225),

(('aufgeteilten', 'divididas'), 17.542956275013225),  
(( 'erwahrt', 'verifitgä'), 17.542956275013225),  
(( 'beschwerdeführer', 'recurrent'), 17.542956275013225),  
(( 'gewissenhaft', 'jau'), 17.542956275013225),  
(( 'gewissenhaft', 'mes'), 17.542956275013225),  
(( 'ich', 'jau'), 17.542956275013225),  
(( 'ich', 'mes'), 17.542956275013225),  
(( 'meines', 'jau'), 17.542956275013225),  
(( 'meines', 'mes'), 17.542956275013225),  
(( 'ratssälen', 'salas'), 17.542956275013225),  
(( '1988', '1988'), 17.542956275013225),  
(( 'parlamentsressourcengesetz', '1988'), 17.542956275013225),  
(( '17a', '17a'), 17.542956275013225),  
(( 'finanzkontrollgesetzes', '1967'), 17.542956275013225),  
(( '1967', '1967'), 17.542956275013225),  
(( 'formulierungen', 'formulaziuns'), 17.542956275013225),  
(( 'sekretariate', 'secretariats'), 17.542956275013225),  
(( 'minderheitsanträge', 'minoritad'), 17.542956275013225),  
(( 'initiant', 'iniziant'), 17.542956275013225),  
(( 'initiantin', 'iniziant'), 17.542956275013225),  
(( 'kommissionsmitglied', 'iniziant'), 17.542956275013225),  
(( 'übernächsten', 'surproxima'), 17.542956275013225),  
(( 'beantwortung', 'respundida'), 17.542956275013225),  
(( 'wahlgängen', 'scrutinis'), 17.542956275013225),  
(( 'indikatoren', 'indicaturs'), 17.542956275013225),  
(( 'administrativuntersuchungen', 'cumenzadas'), 17.542956275013225),  
(( '163-171', '163-171'), 17.542956275013225),  
(( 'funktionieren', 'funcziunament'), 17.542956275013225),  
(( 'übersetzerinnen', 'translaturs'), 17.542956275013225),  
(( 'dezentralisierten', 'decentralisadas'), 17.542956275013225),  
(( 'stimmenmässig', 'dominadas'), 17.542956275013225),  
(( 'beherrscht', 'dominadas'), 17.542956275013225),  
(( '34c', '34c'), 17.542956275013225),  
(( '34b', '34b'), 17.542956275013225),  
(( '36.', '36.'), 17.542956275013225),  
(( '27b', '27b'), 17.542956275013225),  
(( 'suva', 'suva'), 17.542956275013225),  
(( 'auf-', 'ascendent'), 17.542956275013225),  
(( 'bvg', 'lpp'), 17.542956275013225),  
(( '70.', '70.'), 17.542956275013225),  
(( '34a', '34a'), 17.542956275013225),  
(( 'bezeugten', 'attestan'), 17.542956275013225),  
(( 'anmasst', 'usurpescha'), 17.542956275013225),  
(( 'lebendig', 'naschia'), 17.542956275013225),  
(( 'ia', 'ia'), 17.542956275013225),  
(( '50c', '50c'), 17.542956275013225),  
(( 'personenverbindung', 'corporativa'), 17.542956275013225),

(('körperschaftliche', 'corporativa'), 17.542956275013225),  
(('3a', '3a'), 17.542956275013225),  
(('65a', '65a'), 17.542956275013225),  
(('65b', '65b'), 17.542956275013225),  
(('ausländerrechts', 'guntgida'), 17.542956275013225),  
(('inkassohilfe', 'incasso'), 17.542956275013225),  
(('beiseite', 'persistentamain'), 17.542956275013225),  
(('beharrlich', 'persistentamain'), 17.542956275013225),  
(('verständigt', 'approvabla'), 17.542956275013225),  
(('genehmigungsfähigen', 'approvabla'), 17.542956275013225),  
(('aufgehört', 'chalà'), 17.542956275013225),  
(('vermutungen', 'presumziuns'), 17.542956275013225),  
(('gebrechen', 'mendas'), 17.542956275013225),  
(('adoptierenden', 'adoptader'), 17.542956275013225),  
(('adoptionsgeheimnis', 'd`bis'), 17.542956275013225),  
(('reife', 'madirezza'), 17.542956275013225),  
(('gröblich', 'malamain'), 17.542956275013225),  
(('anzehrung', 'consumaziun'), 17.542956275013225),  
(('wiedererlangen', 'recuperaziun'), 17.542956275013225),  
(('selbstbestimmung', 'autodeterminaziun'), 17.542956275013225),  
(('unterlassungen', 'omissiuns'), 17.542956275013225),  
(('vorsorglich', 'preventivamain'), 17.542956275013225),  
(('grossmutter', 'tatta'), 17.542956275013225),  
(('grossmutter', 'tat'), 17.542956275013225),  
(('vorverstorbenen', 'tatta'), 17.542956275013225),  
(('vorverstorbenen', 'tat'), 17.542956275013225),  
(('verschlechterungen', 'pegiuraments'), 17.542956275013225),  
(('vermögensabtretung', 'dota'), 17.542956275013225),  
(('heiratsgut', 'dota'), 17.542956275013225),  
(('glaubt', 'crai'), 17.542956275013225),  
(('gebrechlich', 'mendus'), 17.542956275013225),  
(('abzuwehren', 'dustar'), 17.542956275013225),  
(('ausschliesslicher', 'exclusivs'), 17.542956275013225),  
(('einstimmig', 'unanimamain'), 17.542956275013225),  
(('eigentumsrechts', 'surduvrà'), 17.542956275013225),  
(('wurzeln', 'ragischs'), 17.542956275013225),  
(('winterweg', 'laina'), 17.542956275013225),  
(('winterweg', 'enviern'), 17.542956275013225),  
(('gewässerkorrekturen', 'correcziuns'), 17.542956275013225),  
(('bäche', 'auals'), 17.542956275013225),  
(('abtrete', 'cedia'), 17.542956275013225),  
(('versicherer', 'assicurader'), 17.542956275013225),  
(('bewohnten', 'abitada'), 17.542956275013225),  
(('bewohnten', 'patrun-chasa'), 17.542956275013225),  
(('kostbarer', 'elavuratur'), 17.542956275013225),  
(('verarbeiter', 'elavuratur'), 17.542956275013225),  
(('vermischt', 'maschadadas'), 17.542956275013225),



(('nebensächlich', 'accessoricamain'), 17.542956275013225),  
(('steinbrüchen', 'chavas'), 17.542956275013225),  
(('steinbrüchen', 'crappa'), 17.542956275013225),  
(('rückzahlungen', 'reinvestiziun'), 17.542956275013225),  
(('zwanzigfache', 'ventg'), 17.542956275013225),  
(('verpfändbarkeit', 'impegnabels'), 17.542956275013225),  
(('ursprünglich', 'oriundamain'), 17.542956275013225),  
(('gegenleistungen', 'cuntraprestaziuns'), 17.542956275013225),  
(('xiii', 'xiii'), 17.542956275013225),  
(('aufrufung', 'xiii'), 17.542956275013225),  
(('kaufleuten', 'commerziants'), 17.542956275013225),  
(('teilstück', 'parcella'), 17.542956275013225),  
(('iia', 'iia'), 17.542956275013225),  
(('vermögensmassen', 'massas'), 17.542956275013225),  
(('1972', '1972'), 17.542956275013225),  
(('beglaubigungen', 'legalisaziuns'), 17.542956275013225),  
(('1891', '1891'), 17.542956275013225),  
(('gegenforderungen', 'cuntrapretensiuns'), 17.542956275013225),  
(('08', '08'), 17.542956275013225),  
(('1911', '1911'), 17.542956275013225),  
(('simulation', 'simulaziun'), 17.542956275013225),  
(('vertrauen', 'fidond'), 17.542956275013225),  
(('obligkeitlich', 'concessiunà'), 17.542956275013225),  
(('unbewegliches', 'manglusitad'), 17.542956275013225),  
(('verzinslich', 'tschainsì'), 17.542956275013225),  
(('226f', '226f'), 17.542956275013225),  
(('vorformulierten', 'preformuladas'), 17.542956275013225),  
(('hausbewohnern', 'supportar'), 17.542956275013225),  
(('reinigungen', 'netteggiament'), 17.542956275013225),  
(('termin', 'solvent'), 17.542956275013225),  
(('übersetzten', 'stravagà'), 17.542956275013225),  
(('vertragsänderungen', 'unilateralas'), 17.542956275013225),  
(('mietzinserhöhungen', 'unilateralas'), 17.542956275013225),  
(('1985', '1985'), 17.542956275013225),  
(('sexuell', 'sexualmain'), 17.542956275013225),  
(('belästigt', 'sexualmain'), 17.542956275013225),  
(('mitfinanzierte', 'confinanziada'), 17.542956275013225),  
(('50.', 'accumulada'), 17.542956275013225),  
(('30d-30f', '30d-30f'), 17.542956275013225),  
(('erfindung', 'invenziun'), 17.542956275013225),  
(('design', 'invenziun'), 17.542956275013225),  
(('beeinträchtigt', 'disturbia'), 17.542956275013225),  
(('saison', 'stagiun'), 17.542956275013225),  
(('mangelhaft', 'manglusamain'), 17.542956275013225),  
(('getreue', 'fidaivla'), 17.542956275013225),  
(('einlagerung', 'magasinaziun'), 17.542956275013225),  
(('entwendung', 'destrucziuns'), 17.542956275013225),

(('entwendung', 'accumpagnaders'), 17.542956275013225),  
(( 'entwendung', 'servients'), 17.542956275013225),  
(( 'dienstleute', 'destrucziuns'), 17.542956275013225),  
(( 'dienstleute', 'accumpagnaders'), 17.542956275013225),  
(( 'dienstleute', 'servients'), 17.542956275013225),  
(( 'begleiter', 'destrucziuns'), 17.542956275013225),  
(( 'begleiter', 'accumpagnaders'), 17.542956275013225),  
(( 'begleiter', 'servients'), 17.542956275013225),  
(( 'kostbarkeiten', 'preziusadads'), 17.542956275013225),  
(( 'wagen', 'confidader'), 17.542956275013225),  
(( 'einbringenden', 'confidader'), 17.542956275013225),  
(( 'mitbürgen', 'congarants'), 17.542956275013225),  
(( 'dienstbürgschaften', 'repetan'), 17.542956275013225),  
(( 'vergleiche', 'enclegientschas'), 17.542956275013225),  
(( 'schiedsverträge', 'enclegientschas'), 17.542956275013225),  
(( 'umgekehrt', 'viceversa'), 17.542956275013225),  
(( 'verrechenbarkeit', 'cumpensablada'), 17.542956275013225),  
(( 'hervorgehenden', 'intactas'), 17.542956275013225),  
(( 'gesellschaftssitz', 'exponids'), 17.542956275013225),  
(( 'wiederbeschaffungszwecken', 'reacquisizium'), 17.542956275013225),  
(( 'tantieme', 'tantiema'), 17.542956275013225),  
(( 'unentziehbare', 'nunprivablas'), 17.542956275013225),  
(( 'null', 'nulla'), 17.542956275013225),  
(( 'köpfen', 'chaus'), 17.542956275013225),  
(( 'geschäftszyklus', 'ciclus'), 17.542956275013225),  
(( 'normalen', 'ciclus'), 17.542956275013225),  
(( 'globalurkunden', 'globals'), 17.542956275013225),  
(( 'globalurkunden', 'lubientscha'), 17.542956275013225),  
(( 'gesellschaftsstatuten', 'globals'), 17.542956275013225),  
(( 'gesellschaftsstatuten', 'lubientscha'), 17.542956275013225),  
(( 'ausgestrichene', 'retscha'), 17.542956275013225),  
(( 'ausgestrichene', 'sutsegnà'), 17.542956275013225),  
(( 'stil', 'stil'), 17.542956275013225),  
(( 'zeitberechnung', 'stil'), 17.542956275013225),  
(( 'protesterlass', 'dispensa'), 17.542956275013225),  
(( 'unüberwindliches', 'obstachel'), 17.542956275013225),  
(( 'unüberwindliches', 'surmuntabel'), 17.542956275013225),  
(( 'fortlaufenden', 'numerads'), 17.542956275013225),  
(( 'nummern', 'numerads'), 17.542956275013225),  
(( 'checkfähigkeit', 'passiva'), 17.542956275013225),  
(( 'erdteile', 'continent'), 17.542956275013225),  
(( 'striche', 'stritgs'), 17.542956275013225),  
(( 'indossierbare', 'indossabels'), 17.542956275013225),  
(( '226k', '226k'), 17.542956275013225),  
(( 'fabriken', 'fabricas'), 17.542956275013225),  
(( '1956', '1956'), 17.542956275013225),  
(( '1956', 'impegnativa'), 17.542956275013225),

(('1977', '1977'), 17.542956275013225),  
(( '1918', '1918'), 17.542956275013225),  
(( '1937', '1937'), 17.542956275013225),  
(( 'börsengesetz', 'bursas'), 17.542956275013225),  
(( 'motorfahrzeug-', 'velos'), 17.542956275013225),  
(( 'versicherungspolice', 'polissas'), 17.542956275013225),  
(( 'mediatorin', 'mediatura'), 17.542956275013225),  
(( 'mediator', 'mediatura'), 17.542956275013225),  
(( 'feindschaft', 'amicizia'), 17.542956275013225),  
(( 'feindschaft', 'implitgada'), 17.542956275013225),  
(( 'feindschaft', 'inimicizia'), 17.542956275013225),  
(( 'freundschaft', 'amicizia'), 17.542956275013225),  
(( 'freundschaft', 'implitgada'), 17.542956275013225),  
(( 'freundschaft', 'inimicizia'), 17.542956275013225),  
(( 'eintretende', 'entranta'), 17.542956275013225),  
(( 'gestaltungsklage', 'constitutiv'), 17.542956275013225),  
(( 'zusammengerechnet', 'excludian'), 17.542956275013225),  
(( 'unterliegend', 'perdenta'), 17.542956275013225),  
(( 'befreiungen', 'liberaziuns'), 17.542956275013225),  
(( 'format', 'format'), 17.542956275013225),  
(( 'querulatorische', 'querulantas'), 17.542956275013225),  
(( 'rechtsmissbräuchliche', 'querulantas'), 17.542956275013225),  
(( 'weitschweifige', 'nuncunvegnentas'), 17.542956275013225),  
(( 'weitschweifige', 'illegiblas'), 17.542956275013225),  
(( 'weitschweifige', 'nunchapiblas'), 17.542956275013225),  
(( 'unleserliche', 'nuncunvegnentas'), 17.542956275013225),  
(( 'unleserliche', 'illegiblas'), 17.542956275013225),  
(( 'unleserliche', 'nunchapiblas'), 17.542956275013225),  
(( 'unverständliche', 'nuncunvegnentas'), 17.542956275013225),  
(( 'unverständliche', 'illegiblas'), 17.542956275013225),  
(( 'unverständliche', 'nunchapiblas'), 17.542956275013225),  
(( 'samstag', 'sonda'), 17.542956275013225),  
(( 'revisorinnen', 'faschess'), 17.542956275013225),  
(( 'geheimnisses', 'faschess'), 17.542956275013225),  
(( 'video', 'magnetic'), 17.542956275013225),  
(( 'video', 'bindel'), 17.542956275013225),  
(( 'tonband', 'magnetic'), 17.542956275013225),  
(( 'tonband', 'bindel'), 17.542956275013225),  
(( 'noven', 'novums'), 17.542956275013225),  
(( '68b', '68b'), 17.542956275013225),  
(( 'dispensiert', 'dispenseschia'), 17.542956275013225),  
(( 'erläuterte', 'explitgada'), 17.542956275013225),  
(( 'berichtigte', 'explitgada'), 17.542956275013225),  
(( 'gefährdete', 'sacrifitgar'), 17.542956275013225),  
(( 'preiszugeben', 'sacrifitgar'), 17.542956275013225),  
(( 'reue', 'ricla'), 17.542956275013225),  
(( '189-191', '189-191'), 17.542956275013225),

(('entschuldbaren', 'ferma'), 17.542956275013225),  
(( 'entschuldbaren', 'emoziun'), 17.542956275013225),  
(( 'seelischer', 'ferma'), 17.542956275013225),  
(( 'seelischer', 'emoziun'), 17.542956275013225),  
(( 'gemütsbewegung', 'ferma'), 17.542956275013225),  
(( 'gemütsbewegung', 'emoziun'), 17.542956275013225),  
(( 'heftigen', 'ferma'), 17.542956275013225),  
(( 'heftigen', 'emoziun'), 17.542956275013225),  
(( 'anstaltspfleglingen', 'ricoveradas'), 17.542956275013225),  
(( 'halbbürtigen', 'uterins'), 17.542956275013225),  
(( 'halbbürtigen', 'consanguins'), 17.542956275013225),  
(( 'selbstmord', 'suicidi'), 17.542956275013225),  
(( 'verleitung', 'instigaziun'), 17.542956275013225),  
(( 'feuerwehr', 'pumpiers'), 17.542956275013225),  
(( 'gewinnsucht', 'engurdientscha'), 17.542956275013225),  
(( 'dieb', 'lader'), 17.542956275013225),  
(( 'räuber', 'rapinader'), 17.542956275013225),  
(( 'vorspiegelung', 'vardaivels'), 17.542956275013225),  
(( 'vorspiegelung', 'affirmar'), 17.542956275013225),  
(( 'vorspiegelung', 'supprimer'), 17.542956275013225),  
(( 'vorspiegelung', 'cugliuna'), 17.542956275013225),  
(( 'vorspiegelung', 'affirmescha'), 17.542956275013225),  
(( 'bestärkt', 'vardaivels'), 17.542956275013225),  
(( 'bestärkt', 'affirmar'), 17.542956275013225),  
(( 'bestärkt', 'supprimer'), 17.542956275013225),  
(( 'bestärkt', 'cugliuna'), 17.542956275013225),  
(( 'bestärkt', 'affirmescha'), 17.542956275013225),  
(( 'entschlüsselung', 'decodar'), 17.542956275013225),  
(( 'codierter', 'decodar'), 17.542956275013225),  
(( 'spekulationen', 'speculaziuns'), 17.542956275013225),  
(( 'spekulationen', 'ristgadas'), 17.542956275013225),  
(( 'unehrenhaften', 'suspecta'), 17.542956275013225),  
(( 'unehrenhaften', 'conduita'), 17.542956275013225),  
(( 'unehrenhaften', 'dischonuraivla'), 17.542956275013225),  
(( 'gesprächsteilnehmer', 'conversaziun'), 17.542956275013225),  
(( 'zuführt', 'stauscha'), 17.542956275013225),  
(( 'ausstellungen', 'exposiziuns'), 17.542956275013225),  
(( 'überschwemmung', 'sbuvada'), 17.542956275013225),  
(( 'überschwemmung', 'inundaziun'), 17.542956275013225),  
(( 'schutzvorrichtungen', 'idraulics'), 17.542956275013225),  
(( 'schutzvorrichtungen', 'electrics'), 17.542956275013225),  
(( 'wasserbauten', 'idraulics'), 17.542956275013225),  
(( 'wasserbauten', 'electrics'), 17.542956275013225),  
(( 'pathogene', 'patogens'), 17.542956275013225),  
(( 'unechten', 'autentica'), 17.542956275013225),  
(( 'bescheinigungen', 'attestaziuns'), 17.542956275013225),  
(( 'wasserstandszeichen', 'livel'), 17.542956275013225),

(('vermessungs-', 'livel'), 17.542956275013225),  
(( 'verspottet', 'beffegia'), 17.542956275013225),  
(( 'ethnische', 'etnica'), 17.542956275013225),  
(( 'kombattanten', 'cumbattant'), 17.542956275013225),  
(( 'waffenstillstandes', 'armistizi'), 17.542956275013225),  
(( 'beleidigende', 'vopna'), 17.542956275013225),  
(( 'beleidigende', 'offendents'), 17.542956275013225),  
(( 'fahne', 'vopna'), 17.542956275013225),  
(( 'fahne', 'offendents'), 17.542956275013225),  
(( 'wappen', 'vopna'), 17.542956275013225),  
(( 'wappen', 'offendents'), 17.542956275013225),  
(( 'meuterei', 'revolta'), 17.542956275013225),  
(( 'gütertransportgesetz', '1957'), 17.542956275013225),  
(( 'personenbeförderungsgesetz', '1957'), 17.542956275013225),  
(( '1957', '1957'), 17.542956275013225),  
(( 'eisenbahngesetz', '1957'), 17.542956275013225),  
(( 'sicherheitsorgane', '1957'), 17.542956275013225),  
(( 'bundesamts', '1957'), 17.542956275013225),  
(( 'beleidigung', 'offensiun'), 17.542956275013225),  
(( 'beleidigt', 'radunada'), 17.542956275013225),  
(( 'bekräftigt', 'engirament'), 17.542956275013225),  
(( 'handgelübde', 'engirament'), 17.542956275013225),  
(( 'vorstehers', '14c'), 17.542956275013225),  
(( 'vorstehers', 'lmsi'), 17.542956275013225),  
(( '14c', '14c'), 17.542956275013225),  
(( '14c', 'lmsi'), 17.542956275013225),  
(( 'bwis', '14c'), 17.542956275013225),  
(( 'bwis', 'lmsi'), 17.542956275013225),  
(( 'ausserprozessualen', 'extraprocessuala'), 17.542956275013225),  
(( 'zeugenschutz', 'extraprocessuala'), 17.542956275013225),  
(( 'chiropraktoren', 'chiropratichers'), 17.542956275013225),  
(( 'chiropraktoren', 'psicologs'), 17.542956275013225),  
(( 'patentanwälte', 'chiropratichers'), 17.542956275013225),  
(( 'patentanwälte', 'psicologs'), 17.542956275013225),  
(( 'psychologen', 'chiropratichers'), 17.542956275013225),  
(( 'psychologen', 'psicologs'), 17.542956275013225),  
(( 'eindruck', 'impressiun'), 17.542956275013225),  
(( 'erweckt', 'impressiun'), 17.542956275013225),  
(( '1942', '1942'), 17.542956275013225),  
(( 'verurteilungen', 'condemnaziuns'), 17.542956275013225),  
(( 'archive', 'archivs'), 17.542956275013225),  
(( 'elften', 'indeschavel'), 17.542956275013225),  
(( '31-37', '31-37'), 17.542956275013225),  
(( '31-37', 'delictuala'), 17.542956275013225),  
(( '31-37', 'plausibels'), 17.542956275013225),  
(( 'deliktischen', '31-37'), 17.542956275013225),  
(( 'deliktischen', 'delictuala'), 17.542956275013225),

(('deliktischen', 'plausibels'), 17.542956275013225),  
(('triftige', '31-37'), 17.542956275013225),  
(('triftige', 'delictuala'), 17.542956275013225),  
(('triftige', 'plausibels'), 17.542956275013225),  
(('45.', '45.'), 17.542956275013225),  
(('kollegialgerichten', 'collegialas'), 17.542956275013225),  
(('vormundschaftsbehörden', 'tutelaras'), 17.542956275013225),  
(('strafpunkt', 'culpaivladad'), 17.542956275013225),  
(('bagatellfall', 'bagatella'), 17.542956275013225),  
(('56.', '56.'), 17.542956275013225),  
(('behältnisse', 'recipients'), 17.542956275013225),  
(('körperöffnungen', 'chavorgias'), 17.542956275013225),  
(('körperöffnungen', 'averturas'), 17.542956275013225),  
(('körperhöhlen', 'chavorgias'), 17.542956275013225),  
(('körperhöhlen', 'averturas'), 17.542956275013225),  
(('todesfälle', 'mortoris'), 17.542956275013225),  
(('230bis', '230bis'), 17.542956275013225),  
(('230bis', '264-267'), 17.542956275013225),  
(('230bis', '138-140'), 17.542956275013225),  
(('230bis', '260bis-260quinqies'), 17.542956275013225),  
(('230bis', '322quater'), 17.542956275013225),  
(('230bis', '322ter'), 17.542956275013225),  
(('230bis', '322septies'), 17.542956275013225),  
(('264-267', '230bis'), 17.542956275013225),  
(('264-267', '264-267'), 17.542956275013225),  
(('264-267', '138-140'), 17.542956275013225),  
(('264-267', '260bis-260quinqies'), 17.542956275013225),  
(('264-267', '322quater'), 17.542956275013225),  
(('264-267', '322ter'), 17.542956275013225),  
(('264-267', '322septies'), 17.542956275013225),  
(('260bis-260quinqies', '230bis'), 17.542956275013225),  
(('260bis-260quinqies', '264-267'), 17.542956275013225),  
(('260bis-260quinqies', '138-140'), 17.542956275013225),  
(('260bis-260quinqies', '260bis-260quinqies'), 17.542956275013225),  
(('260bis-260quinqies', '322quater'), 17.542956275013225),  
(('260bis-260quinqies', '322ter'), 17.542956275013225),  
(('260bis-260quinqies', '322septies'), 17.542956275013225),  
(('322quater', '230bis'), 17.542956275013225),  
(('322quater', '264-267'), 17.542956275013225),  
(('322quater', '138-140'), 17.542956275013225),  
(('322quater', '260bis-260quinqies'), 17.542956275013225),  
(('322quater', '322quater'), 17.542956275013225),  
(('322quater', '322ter'), 17.542956275013225),  
(('322quater', '322septies'), 17.542956275013225),  
(('322ter', '230bis'), 17.542956275013225),  
(('322ter', '264-267'), 17.542956275013225),  
(('322ter', '138-140'), 17.542956275013225),

(('322ter', '260bis-260quinquies'), 17.542956275013225),  
(( '322ter', '322quater'), 17.542956275013225),  
(( '322ter', '322ter'), 17.542956275013225),  
(( '322ter', '322septies'), 17.542956275013225),  
(( '322septies', '230bis'), 17.542956275013225),  
(( '322septies', '264-267'), 17.542956275013225),  
(( '322septies', '138-140'), 17.542956275013225),  
(( '322septies', '260bis-260quinquies'), 17.542956275013225),  
(( '322septies', '322quater'), 17.542956275013225),  
(( '322septies', '322ter'), 17.542956275013225),  
(( '322septies', '322septies'), 17.542956275013225),  
(( 'kriegsmaterialgesetz', '34-35b'), 17.542956275013225),  
(( '34-35b', '34-35b'), 17.542956275013225),  
(( 'rechnungsdaten', 'facturaziun'), 17.542956275013225),  
(( 'gewonnene', 'gudagnadas'), 17.542956275013225),  
(( '285a', '285a'), 17.542956275013225),  
(( 'fahnderinnen', 'retschertgaders'), 17.542956275013225),  
(( 'fahnderin', 'retschertgader'), 17.542956275013225),  
(( 'wiederanhandnahme', 'represa'), 17.542956275013225),  
(( 'gerichtsschreiberin', 'actuar'), 17.542956275013225),  
(( '66-73', '66-73'), 17.542956275013225),  
(( 'dänemark', 'danemarc'), 17.542956275013225),  
(( 'mitspracherechte', 'cogestiun'), 17.542956275013225),  
(( 'wettbewerbsverzerrungen', 'distorsions'), 17.542956275013225),  
(( 'verantwortlichkeiten', 'responsabladas'), 17.542956275013225),  
(( 'lehrwerkstätte', 'lavuratori'), 17.542956275013225),  
(( 'tertiärstufe', 'terziar'), 17.542956275013225),  
(( 'methodisch-didaktische', 'spezialitad'), 17.542956275013225),  
(( 'mobilität', 'mobilitad'), 17.542956275013225),  
(( 'stufenweise', 'progressivamain'), 17.542956275013225),  
(( '1800', '1800'), 17.542956275013225),  
(( 'themen', 'temas'), 17.542956275013225),  
(( '1978', '1978'), 17.542956275013225),  
(( 'mieten', 'locaziuns'), 17.542956275013225),  
(( 'rumantsch', 'idioms'), 17.542956275013225),  
(( 'grischun', 'idioms'), 17.542956275013225),  
(( 'schülerinnen', 'scolaras'), 17.542956275013225),  
(( 'schülerinnen', 'scolars'), 17.542956275013225),  
(( 'verlagstätigkeit', 'editoriala'), 17.542956275013225),  
(( 'verlage', 'edituras'), 17.542956275013225),  
(( 'osservatorio', 'osservatorio'), 17.542956275013225),  
(( 'osservatorio', 'linguistico'), 17.542956275013225),  
(( 'osservatorio', 'della'), 17.542956275013225),  
(( 'osservatorio', 'svizzera'), 17.542956275013225),  
(( 'osservatorio', 'italiana'), 17.542956275013225),  
(( 'linguistico', 'osservatorio'), 17.542956275013225),  
(( 'linguistico', 'linguistico'), 17.542956275013225),

(('linguistico', 'della'), 17.542956275013225),  
(( 'linguistico', 'svizzera'), 17.542956275013225),  
(( 'linguistico', 'italiana'), 17.542956275013225),  
(( 'della', 'osservatorio'), 17.542956275013225),  
(( 'della', 'linguistico'), 17.542956275013225),  
(( 'della', 'della'), 17.542956275013225),  
(( 'della', 'svizzera'), 17.542956275013225),  
(( 'della', 'italiana'), 17.542956275013225),  
(( 'svizzera', 'osservatorio'), 17.542956275013225),  
(( 'svizzera', 'linguistico'), 17.542956275013225),  
(( 'svizzera', 'della'), 17.542956275013225),  
(( 'svizzera', 'svizzera'), 17.542956275013225),  
(( 'svizzera', 'italiana'), 17.542956275013225),  
(( 'italiana', 'osservatorio'), 17.542956275013225),  
(( 'italiana', 'linguistico'), 17.542956275013225),  
(( 'italiana', 'della'), 17.542956275013225),  
(( 'italiana', 'svizzera'), 17.542956275013225),  
(( 'italiana', 'italiana'), 17.542956275013225),  
(( 'edk', 'cdep'), 17.542956275013225),  
(( '1966', '1966'), 17.542956275013225),  
(( 'fahrenden', 'viagiants'), 17.542956275013225),  
(( 'stadt', 'citad'), 17.542956275013225),  
(( 'interessenkonflikt', 'durabels'), 17.542956275013225),  
(( 'tresorerie', 'tresoraria'), 17.542956275013225),  
(( 'theater', 'teater'), 17.542956275013225),  
(( 'tanz', 'teater'), 17.542956275013225),  
(( 'sold', 'proviantaziun'), 17.542956275013225),  
(( 'wehrpflichtersatzabgabe', 'exemziun'), 17.542956275013225),  
(( '33-37', '33-37'), 17.542956275013225),  
(( 'kaderkurs', '7-14'), 17.542956275013225),  
(( '7-14', '7-14'), 17.542956275013225),  
(( 'vwvg', 'lfpa'), 17.542956275013225),  
(( 'vstrr', 'dpa'), 17.542956275013225),  
(( 'festsetzungen', 'determinaziuns'), 17.542956275013225),  
(( 'vororientierungen', 'preorientaziuns'), 17.542956275013225),  
(( '1bis', '1bis'), 17.542956275013225),  
(( 'aufstockungen', 'interns'), 17.542956275013225),  
(( '24c', '24c'), 17.542956275013225),  
(( 'genutzte', 'inconfurma'), 17.542956275013225),  
(( 'gebäudevolumens', 'inconfurma'), 17.542956275013225),  
(( 'ethisch', 'eticamain'), 17.542956275013225),  
(( 'vertretbar', 'eticamain'), 17.542956275013225),  
(( 'generationen', 'conscients'), 16.95799377429207),  
(( 'innerrhoden', 'dadens'), 16.95799377429207),  
(( 'innerrhoden', 'appenzell'), 16.95799377429207),  
(( 'innerrhoden', 'sutsilvania'), 16.95799377429207),  
(( 'innerrhoden', 'dador'), 16.95799377429207),



(('innerrhoden', 'sursilvania'), 16.95799377429207),  
(( 'innerrhoden', 'basilea-champagna'), 16.95799377429207),  
(( 'innerrhoden', 'basilea-citad'), 16.95799377429207),  
(( 'basel-stadt', 'dadens'), 16.95799377429207),  
(( 'basel-stadt', 'appenzell'), 16.95799377429207),  
(( 'basel-stadt', 'sutsilvania'), 16.95799377429207),  
(( 'basel-stadt', 'dador'), 16.95799377429207),  
(( 'basel-stadt', 'sursilvania'), 16.95799377429207),  
(( 'basel-stadt', 'basilea-champagna'), 16.95799377429207),  
(( 'basel-stadt', 'basilea-citad'), 16.95799377429207),  
(( 'appenzell', 'dadens'), 16.95799377429207),  
(( 'appenzell', 'appenzell'), 16.95799377429207),  
(( 'appenzell', 'sutsilvania'), 16.95799377429207),  
(( 'appenzell', 'dador'), 16.95799377429207),  
(( 'appenzell', 'sursilvania'), 16.95799377429207),  
(( 'appenzell', 'basilea-champagna'), 16.95799377429207),  
(( 'appenzell', 'basilea-citad'), 16.95799377429207),  
(( 'obwalden', 'dadens'), 16.95799377429207),  
(( 'obwalden', 'appenzell'), 16.95799377429207),  
(( 'obwalden', 'sutsilvania'), 16.95799377429207),  
(( 'obwalden', 'dador'), 16.95799377429207),  
(( 'obwalden', 'sursilvania'), 16.95799377429207),  
(( 'obwalden', 'basilea-champagna'), 16.95799377429207),  
(( 'obwalden', 'basilea-citad'), 16.95799377429207),  
(( 'nidwalden', 'dadens'), 16.95799377429207),  
(( 'nidwalden', 'appenzell'), 16.95799377429207),  
(( 'nidwalden', 'sutsilvania'), 16.95799377429207),  
(( 'nidwalden', 'dador'), 16.95799377429207),  
(( 'nidwalden', 'sursilvania'), 16.95799377429207),  
(( 'nidwalden', 'basilea-champagna'), 16.95799377429207),  
(( 'nidwalden', 'basilea-citad'), 16.95799377429207),  
(( 'freiburg', 'vallais'), 16.95799377429207),  
(( 'freiburg', 'friburg'), 16.95799377429207),  
(( 'ausserrhoden', 'dadens'), 16.95799377429207),  
(( 'ausserrhoden', 'appenzell'), 16.95799377429207),  
(( 'ausserrhoden', 'sutsilvania'), 16.95799377429207),  
(( 'ausserrhoden', 'dador'), 16.95799377429207),  
(( 'ausserrhoden', 'sursilvania'), 16.95799377429207),  
(( 'ausserrhoden', 'basilea-champagna'), 16.95799377429207),  
(( 'ausserrhoden', 'basilea-citad'), 16.95799377429207),  
(( 'basel-landschaft', 'dadens'), 16.95799377429207),  
(( 'basel-landschaft', 'appenzell'), 16.95799377429207),  
(( 'basel-landschaft', 'sutsilvania'), 16.95799377429207),  
(( 'basel-landschaft', 'dador'), 16.95799377429207),  
(( 'basel-landschaft', 'sursilvania'), 16.95799377429207),  
(( 'basel-landschaft', 'basilea-champagna'), 16.95799377429207),  
(( 'basel-landschaft', 'basilea-citad'), 16.95799377429207),

(('wallis', 'friburg'), 16.95799377429207),  
((('zusammenhalt', 'coesiun'), 16.95799377429207),  
((('willkür', 'arbitrariadad'), 16.95799377429207),  
((('grausamer', 'inuman'), 16.95799377429207),  
((('grausamer', 'tortura'), 16.95799377429207),  
((('folter', 'inuman'), 16.95799377429207),  
((('folter', 'crudaivel'), 16.95799377429207),  
((('folter', 'tortura'), 16.95799377429207),  
((('unmenschlicher', 'inuman'), 16.95799377429207),  
((('unmenschlicher', 'tortura'), 16.95799377429207),  
((('glaubens-', 'conscienza'), 16.95799377429207),  
((('glaubens-', 'cretta'), 16.95799377429207),  
((('darbietungen', 'telediffusiun'), 16.95799377429207),  
((('koalitionsfreiheit', 'sindicala'), 16.95799377429207),  
((('streik', 'chauma'), 16.95799377429207),  
((('einbürgerung', 'natiralisaziun'), 16.95799377429207),  
((('finanziell', 'finanzialmain'), 16.95799377429207),  
((('erleiden', 'pateschan'), 16.95799377429207),  
((('gesundheitlichen', 'pateschan'), 16.95799377429207),  
((('breites', 'vast'), 16.95799377429207),  
((('studierende', 'students'), 16.95799377429207),  
((('studierende', 'studentas'), 16.95799377429207),  
((('musik', 'teater'), 16.95799377429207),  
((('angestammten', 'tradiziunalas'), 16.95799377429207),  
((('herkömmliche', 'respectan'), 16.95799377429207),  
((('lästigen', 'nuschaivels'), 16.95799377429207),  
((('verursacher', 'chaschunaders'), 16.95799377429207),  
((('verursacher', 'chaschunadras'), 16.95799377429207),  
((('tier-', 'flora'), 16.95799377429207),  
((('tier-', 'fauna'), 16.95799377429207),  
((('fuss-', 'spassegiar'), 16.95799377429207),  
((('wanderwegnetze', 'spassegiar'), 16.95799377429207),  
((('sparsamen', 'spargnus'), 16.95799377429207),  
((('sparsamen', 'raziunal'), 16.95799377429207),  
((('erneuerbarer', 'spargnus'), 16.95799377429207),  
((('erneuerbarer', 'raziunal'), 16.95799377429207),  
((('erneuerbaren', 'regenerablas'), 16.95799377429207),  
((('erneuerbaren', 'energias'), 16.95799377429207),  
((('brenn-', 'combustibels'), 16.95799377429207),  
((('gestimmt', 'vuschà'), 16.95799377429207),  
((('rasches', 'speditiva'), 16.95799377429207),  
((('wohnungsbau', 'raziunalisaziun'), 16.95799377429207),  
((('anfechtbarkeit', 'contestablidad'), 16.95799377429207),  
((('bundesfeiertag', 'festa'), 16.95799377429207),  
((('krankheiten', 'malsognas'), 16.95799377429207),  
((('keim-', 'schermenza'), 16.95799377429207),  
((('erfahrenen', 'versads'), 16.95799377429207),

(('pubertät', 'pornografics'), 16.95799377429207),  
 (('gleichgewicht', 'equiliber'), 16.95799377429207),  
 (('beherbergungsleistungen', 'alloschament'), 16.95799377429207),  
 (('gegenüberstellen', 'confruntar'), 16.95799377429207),  
 (('versammelt', 'raduna'), 16.95799377429207),  
 (('kollegium', 'collegi'), 16.95799377429207),  
 (('kollegium', 'reunida'), 16.95799377429207),  
 (('truppen', 'truppas'), 16.95799377429207),  
 (('neat', 'vest'), 16.95799377429207),  
 (('neat', 'totalas'), 16.95799377429207),  
 (('mwstg', 'ltp'), 16.95799377429207),  
 (('biersteuer', 'biera'), 16.95799377429207),  
 (('1959', '1959'), 16.95799377429207),  
 (('mitmenschen', 'conumans'), 16.95799377429207),  
 (('übergeordnetem', 'surordinà'), 16.95799377429207),  
 (('variante', 'varianta'), 16.95799377429207),  
 (('unvereinbarkeit', 'incumpatibilitad'), 16.95799377429207),  
 (('regionalverbandes', 'cirquital'), 16.95799377429207),  
 (('umweltschonende', 'ecologic'), 16.95799377429207),  
 (('toleranz', 'toleranza'), 16.95799377429207),  
 (('spezialistinnen', 'spezialistas'), 16.95799377429207),  
 (('qualifizierten', 'digitala'), 16.95799377429207),  
 (('unberechtigtes', 'falsificaziuns'), 16.95799377429207),  
 (('fälschungen', 'falsificaziuns'), 16.95799377429207),  
 (('qualitativ', 'produceschian'), 16.95799377429207),  
 (('ausweisproduktion', 'produceschian'), 16.95799377429207),  
 (('absteigender', 'descendentà'), 16.95799377429207),  
 (('absteigender', 'ascendentà'), 16.95799377429207),  
 (('probleme', 'problems'), 16.95799377429207),  
 (('durchsuchen', 'perquirir'), 16.95799377429207),  
 (('zentren', 'accoglientscha'), 16.95799377429207),  
 (('standortkanton', 'accoglientscha'), 16.95799377429207),  
 (('empfangsstellen', 'accoglientscha'), 16.95799377429207),  
 (('vergewissern', 'persvader'), 16.95799377429207),  
 (('ermessensausübung', 'bainappreziar'), 16.95799377429207),  
 (('immunitäten', 'immunitads'), 16.95799377429207),  
 (('immunitäten', '1961'), 16.95799377429207),  
 (('pass', 'biometric'), 16.95799377429207),  
 (('auslandvertretungen', 'sfalsifitgads'), 16.95799377429207),  
 (('automatisiertes', 'automatisà'), 16.95799377429207),  
 (('teilnehmerkarte', 'biometric'), 16.95799377429207),  
 (('löscht', 'stizza'), 16.95799377429207),  
 (('aliasnamen', 'pseudonims'), 16.95799377429207),  
 (('reisewege', 'itineraris'), 16.95799377429207),  
 (('aufenthaltssorte', 'itineraris'), 16.95799377429207),  
 (('informationsaustausch', '109d'), 16.95799377429207),  
 (('allianz-', 'artist'), 16.95799377429207),

(('kennzeichen', 'artist'), 16.95799377429207),  
(( 'implantate', 'artist'), 16.95799377429207),  
(( 'ordens-', 'artist'), 16.95799377429207),  
(( 'prothesen', 'artist'), 16.95799377429207),  
(( 'diplomatische', '1961'), 16.95799377429207),  
(( 'austauschen', 'barattar'), 16.95799377429207),  
(( '120a', '120a'), 16.95799377429207),  
(( '120b', '120b'), 16.95799377429207),  
(( '120d', '120d'), 16.95799377429207),  
(( 'wiener', '1961'), 16.95799377429207),  
(( 'wiener', 'vienna'), 16.95799377429207),  
(( '1961', 'diplomaticas'), 16.95799377429207),  
(( '1961', 'vienna'), 16.95799377429207),  
(( '1961', 'immunitads'), 16.95799377429207),  
(( 'unvereinbar', 'cumpatiblas'), 16.95799377429207),  
(( '1927', '1927'), 16.95799377429207),  
(( 'komplexe', 'cumplexs'), 16.95799377429207),  
(( 'drucksachen', 'stampats'), 16.95799377429207),  
(( 'übersetzen', 'translateschan'), 16.95799377429207),  
(( 'verzögern', 'retardeschan'), 16.95799377429207),  
(( '1976', '1976'), 16.95799377429207),  
(( 'heimatgemeinde', 'burgaisa'), 16.95799377429207),  
(( 'reste', 'rests'), 16.95799377429207),  
(( 'verwechslungen', 'confusiuns'), 16.95799377429207),  
(( 'bereinigten', 'rectifitgà'), 16.95799377429207),  
(( 'gruppierung', 'gruppaziun'), 16.95799377429207),  
(( 'wählbaren', 'scrutinis'), 16.95799377429207),  
(( 'konstituierende', 'constitutiva'), 16.95799377429207),  
(( 'quorum', 'manchentà'), 16.95799377429207),  
(( 'verfehlt', 'manchentà'), 16.95799377429207),  
(( 'zweifelsfrei', 'irrelevantas'), 16.95799377429207),  
(( 'regierungen', 'regenzas'), 16.95799377429207),  
(( 'sammelfrist', 'rimnar'), 16.95799377429207),  
(( 'vorziehe', 'preferescha'), 16.95799377429207),  
(( 'befangen', 'amicizia'), 16.95799377429207),  
(( 'befangen', 'implitgada'), 16.95799377429207),  
(( 'befangen', 'inimicizia'), 16.95799377429207),  
(( 'parlamenten', 'parlaments'), 16.95799377429207),  
(( '.2', '.2'), 16.95799377429207),  
(( 'planen', 'planiseschan'), 16.95799377429207),  
(( 'vertraulichkeit', 'confidenzialitad'), 16.95799377429207),  
(( 'auszumehren', 'opponidas'), 16.95799377429207),  
(( 'abstimmungsgegenstand', 'opponidas'), 16.95799377429207),  
(( 'vorschlägt', 'solvent'), 16.95799377429207),  
(( 'antwortet', 'respunda'), 16.95799377429207),  
(( 'besetzung', 'vacants'), 16.95799377429207),  
(( 'indikatoren', 'legislativ'), 16.95799377429207),

(('anwendet', 'pressiun'), 16.95799377429207),  
(( 'disziplinaruntersuchungen', 'cumenzadas'), 16.95799377429207),  
(( 'tonträger', 'portatun'), 16.95799377429207),  
(( 'vorsteherin', '14c'), 16.95799377429207),  
(( 'vorsteherin', 'lmsi'), 16.95799377429207),  
(( 'kosd', 'cisl'), 16.95799377429207),  
(( 'kapital-', 'dominadas'), 16.95799377429207),  
(( '663b', '663b'), 16.95799377429207),  
(( 'überzeit', 'sururas'), 16.95799377429207),  
(( 'mehrarbeit', 'sururas'), 16.95799377429207),  
(( '22a', '22a'), 16.95799377429207),  
(( 'efk', 'cdf'), 16.95799377429207),  
(( '27a', '27a'), 16.95799377429207),  
(( 'pensionierung', 'pensiunament'), 16.95799377429207),  
(( 'support', 'support'), 16.95799377429207),  
(( '1^ter', '1^ter'), 16.95799377429207),  
(( 'auf-', 'descendentia'), 16.95799377429207),  
(( '32a', '32a'), 16.95799377429207),  
(( '32d', '32d'), 16.95799377429207),  
(( 'schwankungsreserve', 'fluctuaziun'), 16.95799377429207),  
(( 'beamten-gesetz', '1927'), 16.95799377429207),  
(( 'altem', 'stil'), 16.95799377429207),  
(( 'bezeugten', 'incorrectadad'), 16.95799377429207),  
(( 'unrichtigkeit', 'attestan'), 16.95799377429207),  
(( 'unrichtigkeit', 'incorrectadad'), 16.95799377429207),  
(( 'rechtsgeschäften', 'enclegientschas'), 16.95799377429207),  
(( '1990', '1990'), 16.95799377429207),  
(( 'interessenkonflikte', 'durabels'), 16.95799377429207),  
(( '2-5', '2-5'), 16.95799377429207),  
(( 'sammelvermögen', 'collectada'), 16.95799377429207),  
(( 'beistandschaften', 'curatellas'), 16.95799377429207),  
(( 'ausländerrechts', 'a^bis'), 16.95799377429207),  
(( 'a^bis', 'guntgida'), 16.95799377429207),  
(( 'a^bis', 'a^bis'), 16.95799377429207),  
(( 'bräutigams', 'spusa'), 16.95799377429207),  
(( 'zustimmungen', 'consentiments'), 16.95799377429207),  
(( 'reiflicher', 'reflexiun'), 16.95799377429207),  
(( 'überlegung', 'reflexiun'), 16.95799377429207),  
(( 'überzeugt', 'persvadida'), 16.95799377429207),  
(( 'verständigt', 'accordads'), 16.95799377429207),  
(( 'genehmigungsfähigen', 'accordads'), 16.95799377429207),  
(( 'unterhaltskosten', 'approvabla'), 16.95799377429207),  
(( 'notare', 'chiropratichers'), 16.95799377429207),  
(( 'notare', 'psicologs'), 16.95799377429207),  
(( 'vermögensgegenständen', 'gidà'), 16.95799377429207),  
(( 'kapitalleistung', 'chapitalisada'), 16.95799377429207),  
(( 'zustände', 'chapitalisada'), 16.95799377429207),

(('kapitalwertes', 'chapitalisada'), 16.95799377429207),  
(( 'vermutungen', 'coincidenza'), 16.95799377429207),  
(( '35.', '35avel'), 16.95799377429207),  
(( 'vernachlässigen', 'tralaschan'), 16.95799377429207),  
(( 'sorgfältige', 'conscienziusa'), 16.95799377429207),  
(( 'kombiniert', 'cumbinadas'), 16.95799377429207),  
(( 'apotheker', 'apotechers'), 16.95799377429207),  
(( 'apotheker', 'chiropratichers'), 16.95799377429207),  
(( 'apotheker', 'psicologs'), 16.95799377429207),  
(( 'apothekerinnen', 'apotechers'), 16.95799377429207),  
(( 'zahnärztinnen', 'apotechers'), 16.95799377429207),  
(( 'geistliche', 'chiropratichers'), 16.95799377429207),  
(( 'geistliche', 'psicologs'), 16.95799377429207),  
(( 'wiedererwägung', 'reponderaziun'), 16.95799377429207),  
(( 'befreiende', 'liberant'), 16.95799377429207),  
(( 'graden', 'grads'), 16.95799377429207),  
(( 'graden', 'predefunct'), 16.95799377429207),  
(( 'graden', 'predefuncta'), 16.95799377429207),  
(( 'stämmen', 'grads'), 16.95799377429207),  
(( 'stämmen', 'predefunct'), 16.95799377429207),  
(( 'stämmen', 'predefuncta'), 16.95799377429207),  
(( 'mütterlichen', 'paterna'), 16.95799377429207),  
(( 'mütterlichen', 'materna'), 16.95799377429207),  
(( 'väterlichen', 'paterna'), 16.95799377429207),  
(( 'väterlichen', 'materna'), 16.95799377429207),  
(( 'vorzulesen', 'preleger'), 16.95799377429207),  
(( 'vorgelesen', 'prelegi'), 16.95799377429207),  
(( 'unsittlichkeit', 'immoralitad'), 16.95799377429207),  
(( 'erbunwürdigkeit', 'indignitad'), 16.95799377429207),  
(( 'verheimlicht', 'taschenta'), 16.95799377429207),  
(( 'nachfolgender', 'allontanads'), 16.95799377429207),  
(( 'einwerfung', 'bittar'), 16.95799377429207),  
(( 'nutzungsrechte', 'exclusivs'), 16.95799377429207),  
(( 'versteigert', 'ingiantada'), 16.95799377429207),  
(( 'einstimmigen', 'unanim'), 16.95799377429207),  
(( 'bodenfläche', 'terra'), 16.95799377429207),  
(( 'graben', 'chavar'), 16.95799377429207),  
(( 'nachbarlichen', 'vischinant'), 16.95799377429207),  
(( 'wurzeln', 'roma'), 16.95799377429207),  
(( 'ästen', 'roma'), 16.95799377429207),  
(( 'anries', 'roma'), 16.95799377429207),  
(( 'überbauten', 'surbajegià'), 16.95799377429207),  
(( 'winterweg', 'bavrar'), 16.95799377429207),  
(( 'naturgewalt', 'lavinias'), 16.95799377429207),  
(( 'zufällige', 'lavinias'), 16.95799377429207),  
(( 'bienenschwärme', 'avieuls'), 16.95799377429207),  
(( 'lawinen', 'lavinias'), 16.95799377429207),

(('wind', 'lavinias'), 16.95799377429207),  
(( 'zerstückelung', 'parcellaziun'), 16.95799377429207),  
(( 'güterzusammenlegungen', 'arrundaziuns'), 16.95799377429207),  
(( 'aussehen', 'cumparsa'), 16.95799377429207),  
(( 'umgebung', 'conturns'), 16.95799377429207),  
(( 'nützlichen', 'nizzaivlas'), 16.95799377429207),  
(( 'fahrniseigentums', 'mobigliara'), 16.95799377429207),  
(( 'abnützung', 'isada'), 16.95799377429207),  
(( 'kostspieligen', 'char'), 16.95799377429207),  
(( 'vermischung', 'maschaida'), 16.95799377429207),  
(( 'geschätzte', 'stimadas'), 16.95799377429207),  
(( 'annuität', 'annuitad'), 16.95799377429207),  
(( 'retention', 'cunvegn'), 16.95799377429207),  
(( 'pfandscheines', 'warrant'), 16.95799377429207),  
(( 'warrant', 'warrant'), 16.95799377429207),  
(( 'versäumten', 'racoltar'), 16.95799377429207),  
(( 'einsehbaren', 'chavorgias'), 16.95799377429207),  
(( 'einsehbaren', 'averturas'), 16.95799377429207),  
(( 'löschungen', 'extincziuns'), 16.95799377429207),  
(( 'einwilligen', 'reunids'), 16.95799377429207),  
(( 'zertifikat', 'digitala'), 16.95799377429207),  
(( 'zertifikat', 'certificaziun'), 16.95799377429207),  
(( 'zertifizierungsdiensten', 'digitala'), 16.95799377429207),  
(( 'zertifizierungsdiensten', 'certificaziun'), 16.95799377429207),  
(( '1881', '1881'), 16.95799377429207),  
(( 'gehöre', 'appartegnia'), 16.95799377429207),  
(( 'beweispflichtig', 'appartegnia'), 16.95799377429207),  
(( 'fehlenden', 'mancantas'), 16.95799377429207),  
(( 'zahlungsfähig', 'solvent'), 16.95799377429207),  
(( '50a', '50a'), 16.95799377429207),  
(( 'nachbildung', 'mecanica'), 16.95799377429207),  
(( 'mechanischem', 'mecanica'), 16.95799377429207),  
(( 'schuldbekennntnis', 'fidond'), 16.95799377429207),  
(( 'verbergen', 'zuppentar'), 16.95799377429207),  
(( 'bestattung', 'sepultura'), 16.95799377429207),  
(( 'fortkommens', 'engrevgiament'), 16.95799377429207),  
(( 'leichtes', 'concessiunà'), 16.95799377429207),  
(( 'hilfsperson', 'auxiliara'), 16.95799377429207),  
(( 'übungen', 'isanzas'), 16.95799377429207),  
(( 'mitschuldner', 'cundebitur'), 16.95799377429207),  
(( 'umstand', 'circumstanza'), 16.95799377429207),  
(( 'getränke', 'alcoholicas'), 16.95799377429207),  
(( 'getränke', 'bavrondas'), 16.95799377429207),  
(( 'arglistig', 'malignamain'), 16.95799377429207),  
(( 'arglistig', 'vardaivels'), 16.95799377429207),  
(( 'arglistig', 'affirmar'), 16.95799377429207),  
(( 'arglistig', 'supprimer'), 16.95799377429207),

(('arglistig', 'cugliuna'), 16.95799377429207),  
(( 'arglistig', 'affirmescha'), 16.95799377429207),  
(( 'minderungsklage', 'redebitaziun'), 16.95799377429207),  
(( 'wandlungsklage', 'redebitaziun'), 16.95799377429207),  
(( 'prozenten', 'pertschients'), 16.95799377429207),  
(( 'kapitalzins', 'capitada'), 16.95799377429207),  
(( 'mietsache', 'stravagà'), 16.95799377429207),  
(( 'geschäftsbedingungen', 'preformuladas'), 16.95799377429207),  
(( 'rückgabeprotokoll', 'vesair'), 16.95799377429207),  
(( 'besichtigen', 'relocaziun'), 16.95799377429207),  
(( 'besichtigen', 'inspectar'), 16.95799377429207),  
(( '259a-259i', '259a-259i'), 16.95799377429207),  
(( 'untermieter', 'sutlocatari'), 16.95799377429207),  
(( 'konsumentenpreise', 'index'), 16.95799377429207),  
(( 'landesindex', 'index'), 16.95799377429207),  
(( 'gestaffelte', 'graduads'), 16.95799377429207),  
(( 'viehpacht', 'affittaziun'), 16.95799377429207),  
(( 'viehverstellung', 'affittaziun'), 16.95799377429207),  
(( 'verschlechtert', 'pegiuradas'), 16.95799377429207),  
(( 'pfändbar', 'impegnablas'), 16.95799377429207),  
(( 'auszurüsten', 'equipar'), 16.95799377429207),  
(( 'gebrauchs', 'bunifitgadas'), 16.95799377429207),  
(( 'wohnbaugenossenschaft', 'confinanziada'), 16.95799377429207),  
(( 'vollmachten', 'plainpudairs'), 16.95799377429207),  
(( 'literarischen', 'litterara'), 16.95799377429207),  
(( 'zeitschriften', 'gasettas'), 16.95799377429207),  
(( 'zeitschriften', 'revistas'), 16.95799377429207),  
(( 'partnerschaftsvermittlung', 'maridaglias'), 16.95799377429207),  
(( 'adressat', 'consegnader'), 16.95799377429207),  
(( 'unwiderruflich', 'irrevocabla'), 16.95799377429207),  
(( 'entwendung', 'cumprovian'), 16.95799377429207),  
(( 'entwendung', 'visitaders'), 16.95799377429207),  
(( 'besucher', 'destrucziuns'), 16.95799377429207),  
(( 'besucher', 'accompagnaders'), 16.95799377429207),  
(( 'besucher', 'servients'), 16.95799377429207),  
(( 'besucher', 'visitaders'), 16.95799377429207),  
(( 'dienstleute', 'cumprovian'), 16.95799377429207),  
(( 'dienstleute', 'visitaders'), 16.95799377429207),  
(( 'begleiter', 'cumprovian'), 16.95799377429207),  
(( 'begleiter', 'visitaders'), 16.95799377429207),  
(( 'dienstbürgschaften', 'vitgira'), 16.95799377429207),  
(( 'rechtsweg', 'interrupziuns'), 16.95799377429207),  
(( 'lotterie-', 'lottarias'), 16.95799377429207),  
(( 'soliden', 'solida'), 16.95799377429207),  
(( 'rentenanstalt', 'solida'), 16.95799377429207),  
(( 'hindernis', 'obstachel'), 16.95799377429207),  
(( 'hindernis', 'surmuntabel'), 16.95799377429207),



```
(('kommanditäre', 'commanditaris'), 16.95799377429207),
(('einladung', 'invit'), 16.95799377429207),
(('unsachlicher', 'dischavantagià'), 16.95799377429207),
(('unsachlicher', 'favurisà'), 16.95799377429207),
(('säumigen', 'perdia'), 16.95799377429207),
(('verhandlungsgegenstandes', 'tractanda'), 16.95799377429207),
(('objektiv', 'objectiva'), 16.95799377429207),
(('anschein', 'objectiva'), 16.95799377429207),
(('prüfungsurteil', 'objectiva'), 16.95799377429207),
(('zusammenfassenden', 'resumescha'), 16.95799377429207),
(('sachgemässe', 'supportabla'), 16.95799377429207),
(('bestellscheinen', 'empustaziun'), 16.95799377429207),
(('beteiligungspapiere', 'quotads'), 16.95799377429207)]
```

Inspect the highest MI scores for «ergebnis»:

```
[16]: sorted({k: v for k, v in ranking.items() if k[0] == 'ergebnis'}.items(),
→key=lambda item: item[1], reverse=True)[:10]
```

```
[16]: [(('ergebnis', 'verifitgà'), 13.413673258068261),
(('ergebnis', 'resumescha'), 13.413673258068261),
(('ergebnis', 'resultat'), 12.323475449096682),
(('ergebnis', 'rimnada'), 12.091745163180898),
(('ergebnis', 'indirecta'), 11.413673258068261),
(('ergebnis', 'falsifitgescha'), 11.298196040648325),
(('ergebnis', 'constataziuns'), 10.32621041681792),
(('ergebnis', 'total'), 9.954241639430963),
(('ergebnis', 'cuntraproposta'), 9.954241639430963),
(('ergebnis', 'suttametta'), 9.921820161738585)]
```

Inspect the highest MI scores for words starting with «viagiant»:

```
[17]: sorted({k: v for k, v in ranking.items() if k[1].startswith('viagiant')}.
→items(), key=lambda item: item[1], reverse=True)[:10]
```

```
[17]: [(('fahrenden', 'viagiants'), 17.542956275013225),
(('fahrenden', 'viagiantas'), 16.95799377429207),
(('ihrer', 'viagiantas'), 9.112503723347695),
(('der', 'viagiantas'), 5.182657848395181),
(('der', 'viagiants'), 5.182657848395181),
(('', 'viagiantas'), 4.7100662608484845),
(('.', 'viagiantas'), 4.122634152483831)]
```