

```
import re
import os
```

Make sure that `hfst-lookup bezhta.tr.analyzer.hfst` is in the same directory

```
def check_word(word):
    output = os.popen(f"echo {word} | hfst-lookup
bezhta.tr.analyzer.hfst").read().rstrip('\n\n')
    parses = []
    for el in output.split('\n'):
        parses.append(':'.join(el.split('\t')))

    return parses

with open('valpal_examples.txt', 'r', encoding='utf-8') as file:
    file = file.readlines()

from collections import defaultdict

def check_tr_coverage(path):
    with open(path, 'r', encoding='utf-8') as file:
        un = []
        total = 0.0
        tokens_recognised = 0.0
        unique = 0.0
        types_recognised = 0.0
        unrecog = defaultdict(int)
        lines = file.readlines()
        for l in lines:
            w = l.strip()
            parses = check_word(w)
            if '?' not in parses[0]:
                tokens_recognised += 1
                if w not in un:
                    types_recognised += 1
            else:
                unrecog[w] += 1
            if w not in un:
                un.append(w)
            total += 1
        print('tokens: ', total)
        print('types: ', len(un))
        print('tokens recognised: ', tokens_recognised)
        print('token coverage: ', tokens_recognised / total)
        print('types recognised: ', types_recognised)
        print('type coverage: ', types_recognised / len(un))
        return unrecog

check_tr_coverage('valpal_examples.txt')
```

tokens: 1137.0
types: 462
tokens recognised: 811.0
token coverage: 0.7132805628847845
types recognised: 256.0
type coverage: 0.5541125541125541

```
defaultdict(int,  
             {'hič'egolca': 1,  
              'łic'o': 3,  
              'iXeš': 1,  
              'yiyas': 1,  
              'dilna': 2,  
              'uštelli': 2,  
              'meeyo': 2,  
              'xöxł'äš': 4,  
              'xöxił': 4,  
              'łi': 7,  
              'istolijał'a': 5,  
              't'ek': 24,  
              'mašinalł'a': 6,  
              'isilco': 3,  
              'urxeziaqiyo': 1,  
              'xalbooyo': 1,  
              'goc'oqiyo': 1,  
              'mähmädi': 2,  
              'murad': 3,  
              'kuntalałco': 2,  
              'pacpałos': 1,  
              'łiis': 3,  
              'łanq'izibaqiyo': 1,  
              'äłał'aš': 3,  
              'č'an': 3,  
              'guus': 1,  
              'ałto': 3,  
              'kusułiyo': 1,  
              'keziyaqiyo': 1,  
              'simindiyala': 3,  
              'hät't'äš': 3,  
              'ist'i': 2,  
              'okko': 3,  
              'harziyooyo': 1,  
              'bišello': 2,  
              'xöxł'ä': 3,  
              'yurłan': 2,  
              'laladayo': 1,  
              't'eklad': 4,  
              'öhdäyo': 1,  
              'ye"eyo': 3,  
              'nänilał': 2,
```

'łowas': 1,
'q'ac'icco': 4,
'biłoqan': 1,
'yiqoyo': 1,
'zistuqilca': 1,
'łiid': 2,
'goldas': 2,
'q'ac'iccad': 1,
'istoliyał'as': 2,
'aⁿcoldas': 1,
'bäbäläd': 4,
'üč'das': 1,
'uștel': 1,
'eⁿhdas': 2,
'łic'olid': 1,
'iłelas': 1,
'bek'eldas': 1,
'yokčildas': 1,
'haboł'a': 4,
'mehca': 1,
'k'äyłölca': 1,
'urxeziyooyo': 1,
'gič'eqilca': 1,
'iyagolca': 1,
'pacpałolca': 1,
'?arab': 3,
'małciboos': 1,
'č'anaxoš': 1,
'k'isagolca': 1,
'öždää': 1,
'kezibooyo': 1,
'boylis': 1,
'ist'id': 18,
'yoyliyo': 1,
'yeⁿeyliyo': 1,
'yežeylis': 1,
'yurıanlid': 1,
'yoq'oldas': 1,
'łowagilca': 1,
'micał': 1,
'äkid': 1,
'äxdäš': 1,
'bäxillis': 1,
'yiqollo': 1,
'pacpałoyo': 1,
'yüⁿqdagilca': 1,
'nizayliyo': 1,
'iłellis': 1,
'boyliyo': 1,

'nisoyliyo': 1,
'meeyliyo': 1,
'maʽcibooyliyo': 1,
'gulliyo': 1,
'urʽezibooyo': 1,
'is': 1,
'goc'oqilliyo': 1,
'nukogolca': 1,
'diʽilād': 1,
'xalboyliyo': 1,
'ʽiç'eldas': 1,
'māhmād': 1,
'kuntalaʽla': 1,
'aʽʽod': 1,
'oc'ildas': 1,
'yeʽnyliyo': 1,
'ʽiç'elliyo': 1,
'yāč'k'iyo': 1,
'öhdāgilca': 1,
'guyliyo': 1,
'harziyoyliyo': 1,
'müqiyʽliyo': 1,
'nizayʽliyo': 1,
'iʽkeyʽliyo': 1,
'mosoyʽliyo': 1,
'micoyʽliyo': 1,
'git'iyʽliyo': 1,
'zistuqilliyo': 1,
'biʽʽeyʽliyo': 2,
'meyliyo': 1,
'baq'oyʽliyo': 1,
'yuʽncoyliyo': 1,
'yeʽzeyʽliyo': 1,
'eʽneyo': 1,
'yeʽnyʽliyo': 1,
'niʽʽiyʽliyo': 1,
'ʽanq'izibooyo': 1,
'yuʽncoyʽliyo': 2,
'booyliyo': 1,
'woyid': 1,
'hič'egolilliyo': 1,
'büč'iyʽliyo': 1,
'haʽoyliyo': 1,
'bāʽʽeyʽliyo': 1,
'nisoyʽliyo': 1,
'uʽtelliqa': 1,
'meeyʽliyo': 1,
'booyʽliyo': 2,
'pacpaʽoyʽliyo': 1,

'yokčiyłiyo': 1,
'č'anaxoyłiyo': 1,
'małcibooyłiyo': 1,
'mähmädqa': 1,
'guliylıyo': 2,
'haʔoyłiyo': 1,
'bäxiylıyo': 1,
'xalboyłiyo': 1,
'ist'ıqa': 1,
'harziyooyłiyo': 1,
'yoc'ıylıyo': 1,
'öhxöyłiyo': 1,
'lalaʁoyłiyo': 1,
'kusuʔıylıyo': 1,
'yiyayłiyo': 2,
'k'ok'oyłiyo': 1,
'qoqoyłiyo': 1,
'boyłiyo': 1,
'bek'eyłiyo': 1,
'bekiyłiyo': 1,
'xanq'izibaqiylıyo': 1,
'xayʁoyłiyo': 1,
'k'äyʁöyłiyo': 1,
'heleyłiyo': 1,
'goc'oqiylıyo': 1,
'eⁿx'eyłiyo': 3,
'yečeyłiyo': 1,
'k'isayłiyo': 1,
'maʔiřatlis': 1,
'urxeziaqiylıyo': 1,
'yiq'lařeř': 1,
'yiq'lařeylıyo': 2,
'ogic'ıylıyo': 1,
'yogic'ıylıyo': 1,
'begaayłiyo': 1,
'vedra': 1,
'istolıyaʁ'a': 1,
'hič'eyłiyo': 1,
'biq'eyłiyo': 1,
'nukoyłiyo': 1,
'yoⁿq'oyłiyo': 1,
'öždiqa': 1,
'mexellis': 1,
'yaʁoyłiyo': 1,
'č'anaxoylis': 1,
'yaʁoylis': 1,
'iq'lařegolca': 1,
'łič'eyłiyo': 1,
'iřeldas': 1,

'č'anad': 1,
'aɣoldas': 1,
'k'iyasas': 1,
'pacpaɣoldas': 1,
'yegaaldas': 1,
'tohdas': 1,
'zistuqildas': 1,
'yiq'eldas': 1,
'fenlid': 1,
'müč'lid': 1,
'qoqoldas': 1,
'ɣayɣoldas': 1,
'yoc'ildas': 1,
'eⁿšlid': 1,
'yekildas': 1,
'yohdas': 1,
'ɭič'eyo': 1})