PREDICTION OF COGNATE REFLEXES

TEAM MEMBERS:
Varshaa Shree Bhuvanendar (G01269710)
Ishana Vikram Shinde (G01268126)
Deeksha Gangadharan Srinivas (G01291097)
Kavya Sudha Kollu (G01272848)

<u>Introduction</u>:

Cognates share a common origin regardless of their meaning and don’t contain borrowed words. Individual members in the cognate set also known as cognate reflex show similar sound patterns with other members of the cognate set. This allows mapping across the individual phoneme systems of the individual languages where most of the time the mappings depend on contextual conditions that differ based on their positions in the word. By leveraging this we develop an approach to predict Cognate Reflexes using SOTA techniques.

<u>Analysis</u>:

The best results were achieved by the Mockingbird team where they used two models -
1. The Neighbor Transformer model. This model extends transformer-based encode-decoder sequence-to-sequence modelling, by encoding all available input cognates in parallel and having the decoder attend to the resulting joint representation during inference.

2. Image Inpainting Model - This model compares the cognate reflex prediction task to the task of restoring corrupted parts of a 2D image, in which dimensions correspond to languages and cognate phonemic representations. The restoration is achieved with the help of convolutional neural networks.

We have tried to replicate the results of the Image Inpainting Model for our baseline.
We tried to implement the paper (mockingbird) that gave the best results, and we were able to get similar results on the training datasets.

<u>Experiments:</u>

- We tried on data of proportion 0.10 and 0.50
- Tried on surprise cognate set of languages and provided train set of cognate languages

NOTE: The paper did a similar experiment.

<u>Error Analysis:</u>

We noticed that for surprise data i.e., data for which we did not have any dev data, we changed the logic a bit for such models the checkpoint file gets created over all epochs, and for the train, we pass the best checkpoint file. It was observed that this leads to a drastic change in BLUE Score mainly that the best checkpoint file is not getting picked rather the last ran epoch model is taken for the test.

<u>Results</u>:

How does your reproduced model compare with the reported SOTA?

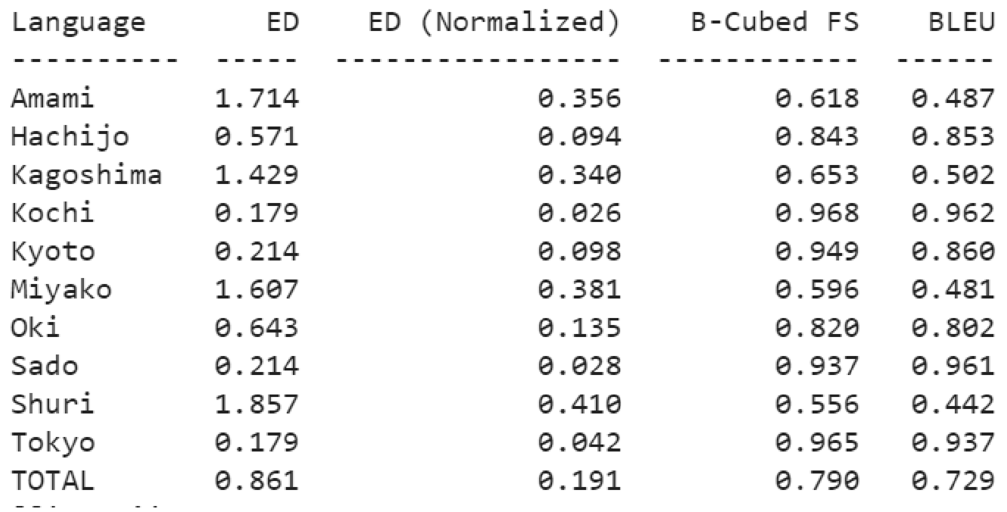
- The reproduced model is close to the SOTA Base model
- There is a difference of 0.2 -0.5 numerical value for BLEU Score

Link to Dataset: https://zenodo.org/record/6567339#.Y2Xa33bMK3A

<u>RESULTS for the split 0.10</u>

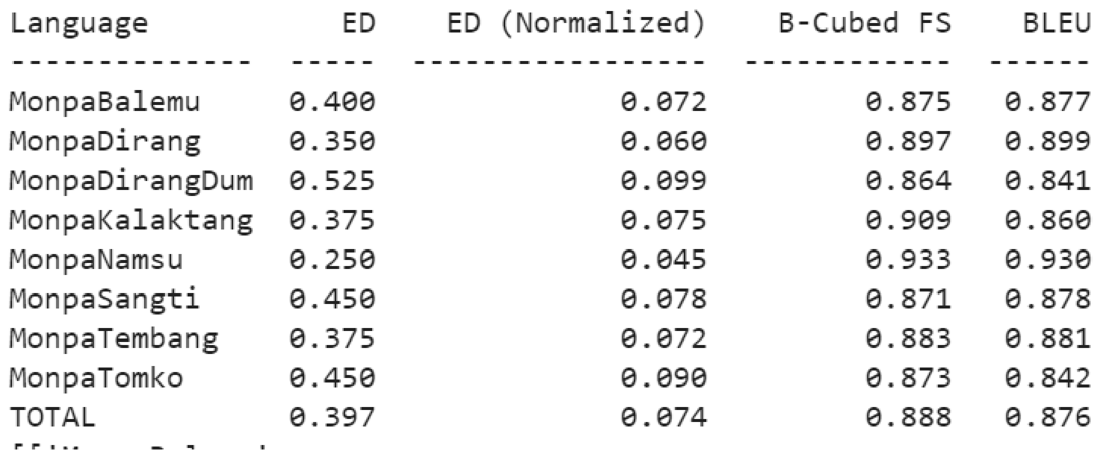
<u>TRAINING DATA SET</u>

1.Dataset: hattorijaponic



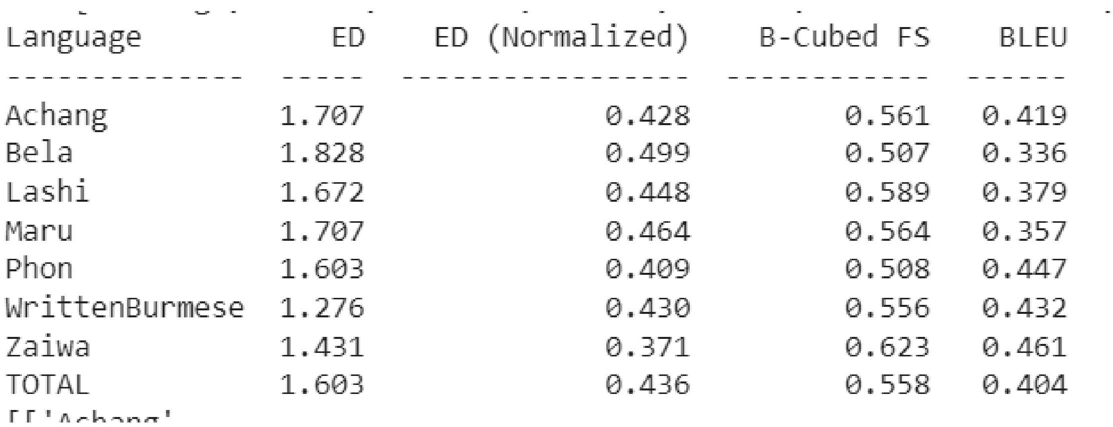
```
Language       ED     ED (Normalized)    B-Cubed FS    BLEU
----------    -----  -----------------  ------------  ------
Amami         1.714              0.356         0.618   0.487
Hachijo       0.571              0.094         0.843   0.853
Kagoshima     1.429              0.340         0.653   0.502
Kochi         0.179              0.026         0.968   0.962
Kyoto         0.214              0.098         0.949   0.860
Miyako        1.607              0.381         0.596   0.481
Oki           0.643              0.135         0.820   0.802
Sado          0.214              0.028         0.937   0.961
Shuri         1.857              0.410         0.556   0.442
Tokyo         0.179              0.042         0.965   0.937
TOTAL         0.861              0.191         0.790   0.729
```

2.Dataset: abrahammonpa



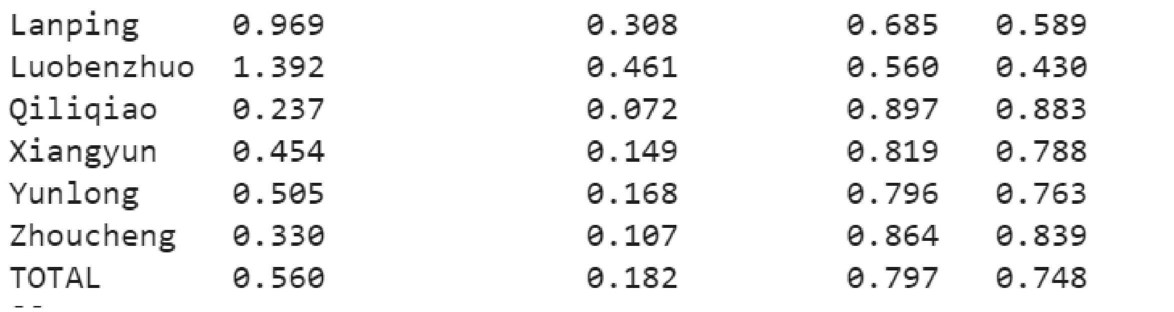
```
Language            ED     ED (Normalized)    B-Cubed FS    BLEU
--------------    -----  -----------------  ------------  ------
MonpaBalemu       0.400              0.072         0.875   0.877
MonpaDirang       0.350              0.060         0.897   0.899
MonpaDirangDum    0.525              0.099         0.864   0.841
MonpaKalaktang    0.375              0.075         0.909   0.860
MonpaNamsu        0.250              0.045         0.933   0.930
MonpaSangti       0.450              0.078         0.871   0.878
MonpaTembang      0.375              0.072         0.883   0.881
MonpaTomko        0.450              0.090         0.873   0.842
TOTAL             0.397              0.074         0.888   0.876
```

3.Dataset: Manburmish



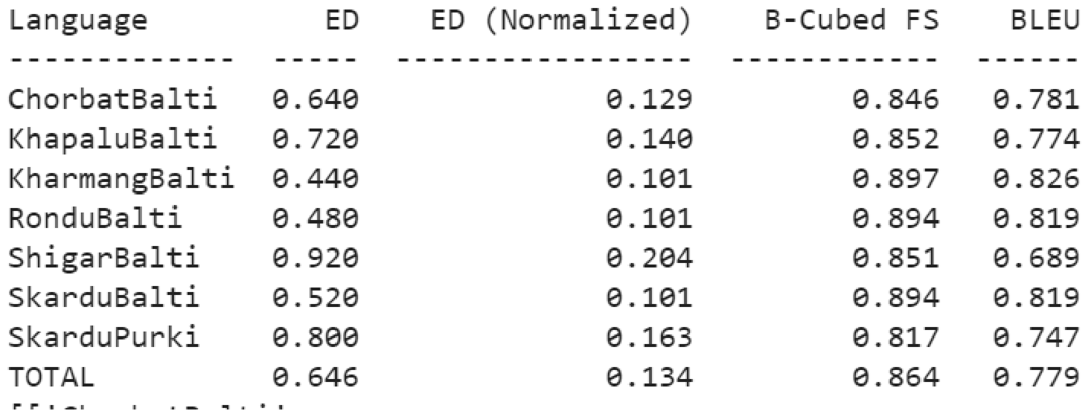
```
Language           ED    ED (Normalized)    B-Cubed FS    BLEU
--------------   -----  -----------------  ------------  ------
Achang           1.707              0.428         0.561   0.419
Bela             1.828              0.499         0.507   0.336
Lashi            1.672              0.448         0.589   0.379
Maru             1.707              0.464         0.564   0.357
Phon             1.603              0.409         0.508   0.447
WrittenBurmese   1.276              0.430         0.556   0.432
Zaiwa            1.431              0.371         0.623   0.461
TOTAL            1.603              0.436         0.558   0.404
```

4.Dataset: allenbai



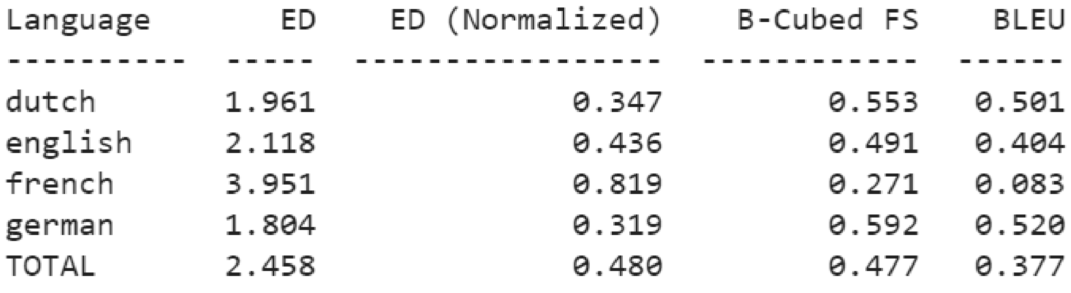
```
Lanping      0.969              0.308         0.685   0.589
Luobenzhuo   1.392              0.461         0.560   0.430
Qiliqiao     0.237              0.072         0.897   0.883
Xiangyun     0.454              0.149         0.819   0.788
Yunlong      0.505              0.168         0.796   0.763
Zhoucheng    0.330              0.107         0.864   0.839
TOTAL        0.560              0.182         0.797   0.748
```

5.Dataset: backstromnorthernpakistan



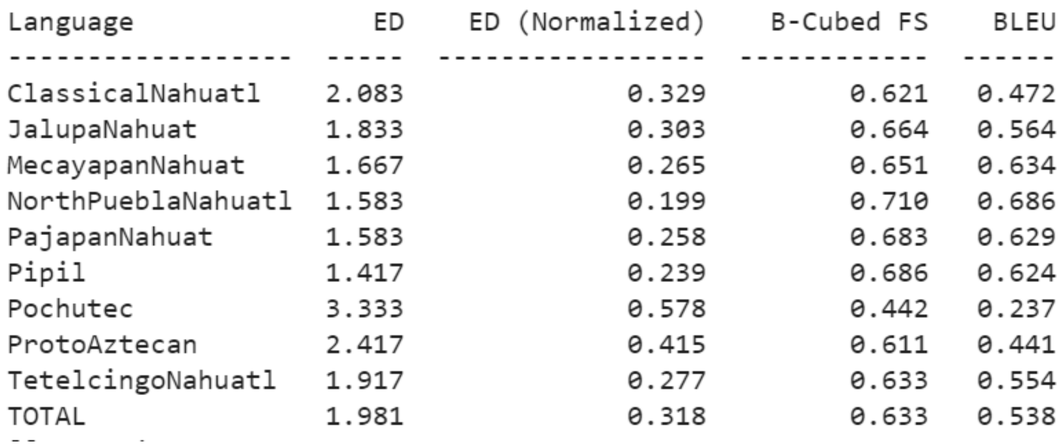
```
Language          ED    ED (Normalized)    B-Cubed FS    BLEU
-------------   -----  -----------------  ------------  ------
ChorbatBalti    0.640              0.129         0.846   0.781
KhapaluBalti    0.720              0.140         0.852   0.774
KharmangBalti   0.440              0.101         0.897   0.826
RonduBalti      0.480              0.101         0.894   0.819
ShigarBalti     0.920              0.204         0.851   0.689
SkarduBalti     0.520              0.101         0.894   0.819
SkarduPurki     0.800              0.163         0.817   0.747
TOTAL           0.646              0.134         0.864   0.779
```

6.Dataset: Listsamplesize



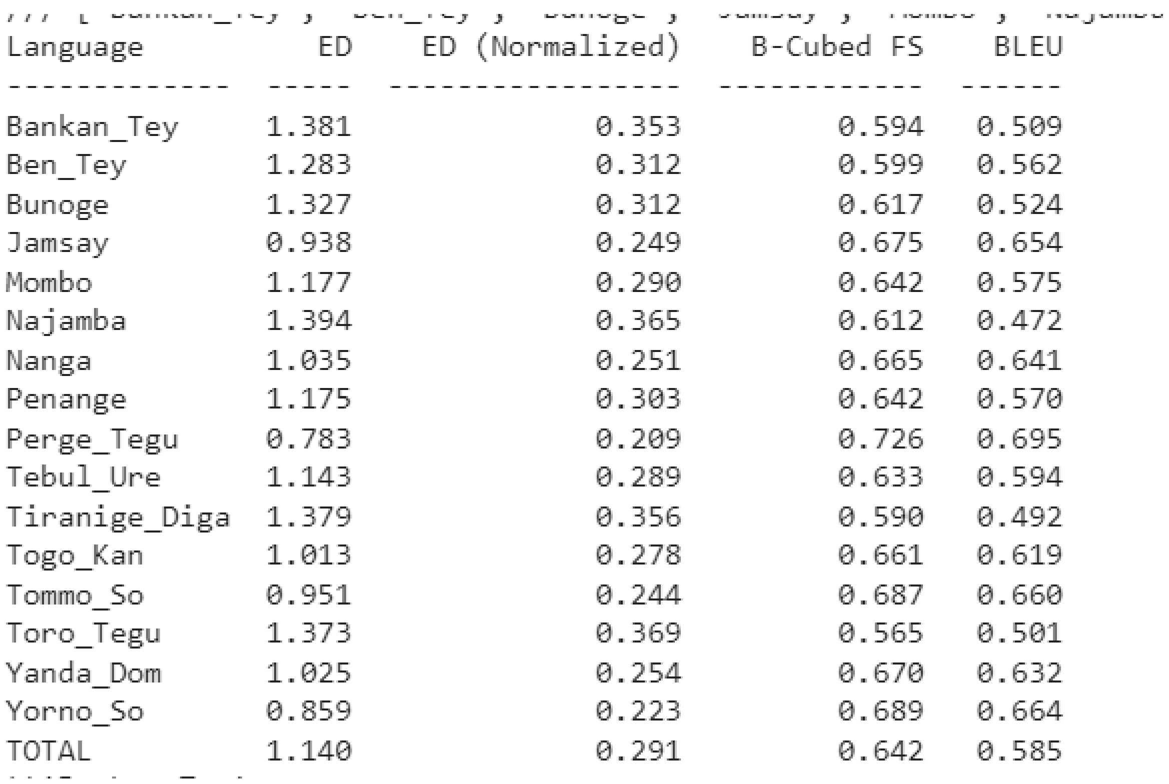
```
Language       ED    ED (Normalized)    B-Cubed FS    BLEU
----------   -----  -----------------  ------------  ------
dutch        1.961              0.347         0.553   0.501
english      2.118              0.436         0.491   0.404
french       3.951              0.819         0.271   0.083
german       1.804              0.319         0.592   0.520
TOTAL        2.458              0.480         0.477   0.377
```

7. Language: davletshinaztecan



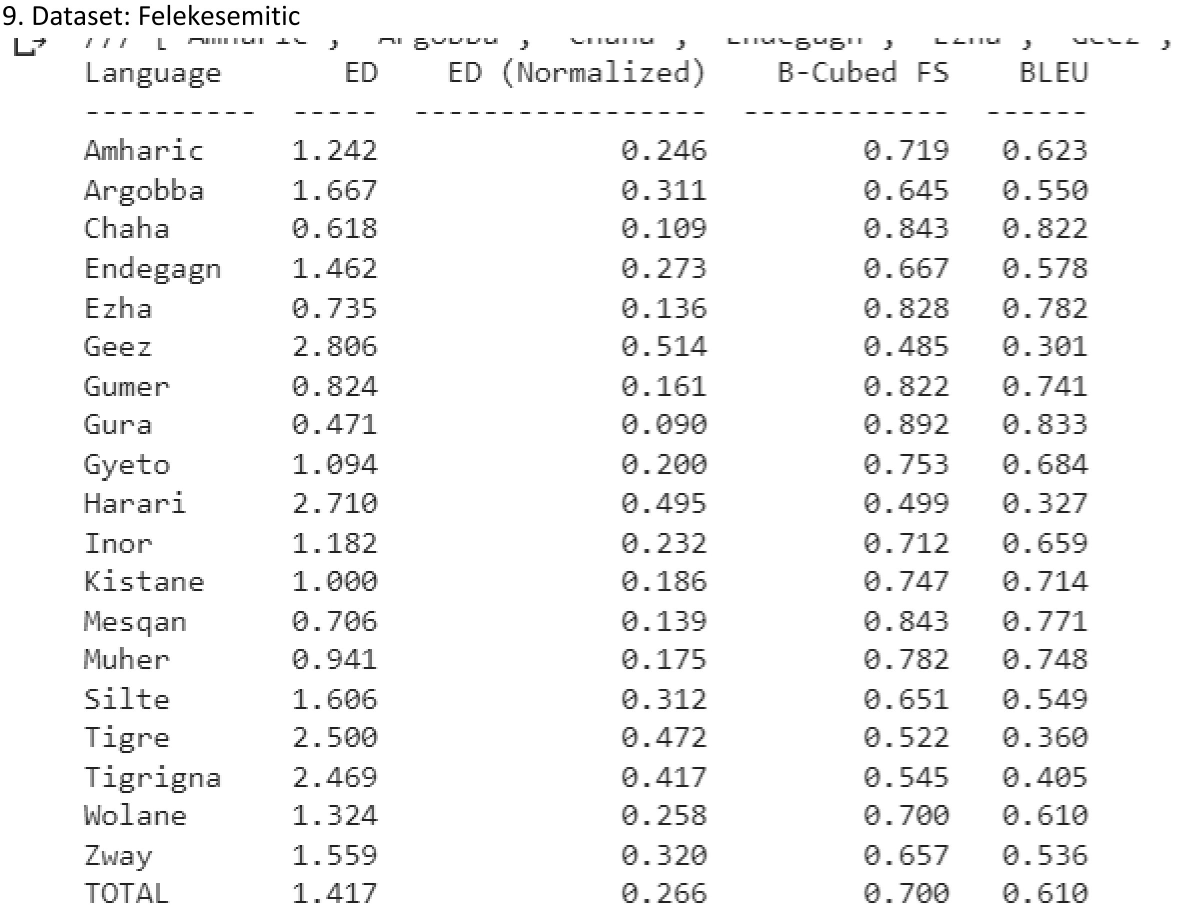
```
Language               ED    ED (Normalized)    B-Cubed FS    BLEU
------------------  -----  -----------------  ------------  ------
ClassicalNahuatl    2.083              0.329         0.621   0.472
JalupaNahuat        1.833              0.303         0.664   0.564
MecayapanNahuat     1.667              0.265         0.651   0.634
NorthPueblaNahuatl  1.583              0.199         0.710   0.686
PajapanNahuat       1.583              0.258         0.683   0.629
Pipil               1.417              0.239         0.686   0.624
Pochutec            3.333              0.578         0.442   0.237
ProtoAztecan        2.417              0.415         0.611   0.441
TetelcingoNahuatl   1.917              0.277         0.633   0.554
TOTAL               1.981              0.318         0.633   0.538
```

8.Dataset: Hantganbangime



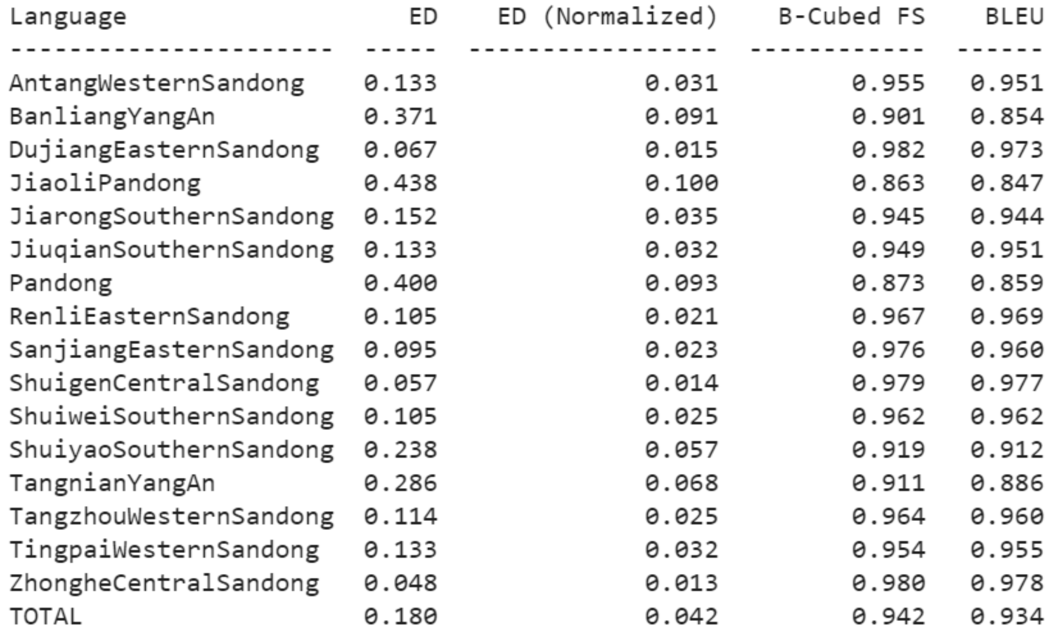
```
Language          ED    ED (Normalized)    B-Cubed FS    BLEU
-------------  -----  -----------------  ------------  ------
Bankan_Tey     1.381              0.353         0.594   0.509
Ben_Tey        1.283              0.312         0.599   0.562
Bunoge         1.327              0.312         0.617   0.524
Jamsay         0.938              0.249         0.675   0.654
Mombo          1.177              0.290         0.642   0.575
Najamba        1.394              0.365         0.612   0.472
Nanga          1.035              0.251         0.665   0.641
Penange        1.175              0.303         0.642   0.570
Perge_Tegu     0.783              0.209         0.726   0.695
Tebul_Ure      1.143              0.289         0.633   0.594
Tiranige_Diga  1.379              0.356         0.590   0.492
Togo_Kan       1.013              0.278         0.661   0.619
Tommo_So       0.951              0.244         0.687   0.660
Toro_Tegu      1.373              0.369         0.565   0.501
Yanda_Dom      1.025              0.254         0.670   0.632
Yorno_So       0.859              0.223         0.689   0.664
TOTAL          1.140              0.291         0.642   0.585
```

9. Dataset: Felekesemitic



```
Language       ED    ED (Normalized)    B-Cubed FS    BLEU
----------  -----  -----------------  ------------  ------
Amharic     1.242              0.246         0.719   0.623
Argobba     1.667              0.311         0.645   0.550
Chaha       0.618              0.109         0.843   0.822
Endegagn    1.462              0.273         0.667   0.578
Ezha        0.735              0.136         0.828   0.782
Geez        2.806              0.514         0.485   0.301
Gumer       0.824              0.161         0.822   0.741
Gura        0.471              0.090         0.892   0.833
Gyeto       1.094              0.200         0.753   0.684
Harari      2.710              0.495         0.499   0.327
Inor        1.182              0.232         0.712   0.659
Kistane     1.000              0.186         0.747   0.714
Mesqan      0.706              0.139         0.843   0.771
Muher       0.941              0.175         0.782   0.748
Silte       1.606              0.312         0.651   0.549
Tigre       2.500              0.472         0.522   0.360
Tigrigna    2.469              0.417         0.545   0.405
Wolane      1.324              0.258         0.700   0.610
Zway        1.559              0.320         0.657   0.536
TOTAL       1.417              0.266         0.700   0.610
```

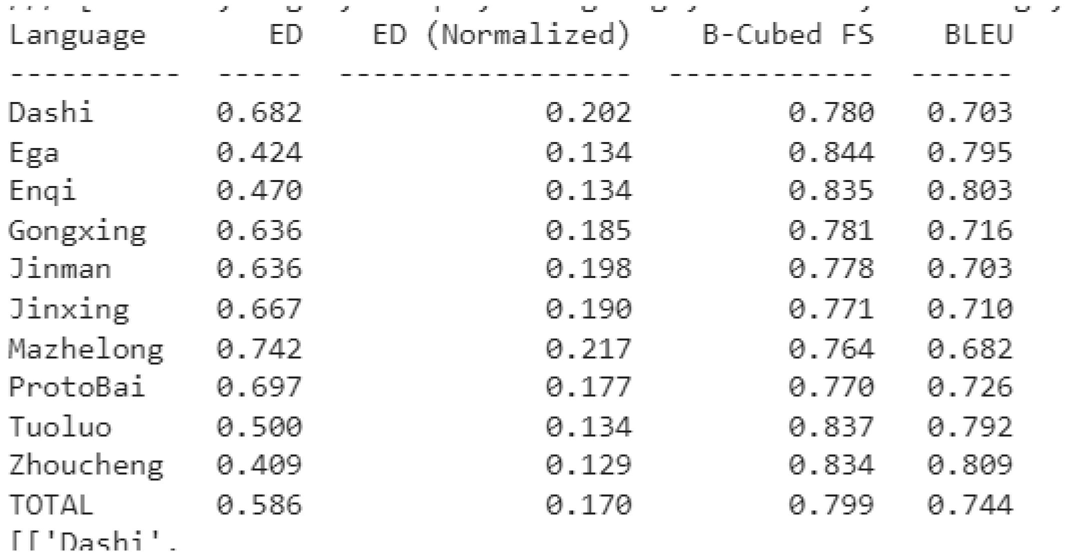
10. Dataset: castrosui



```
Language                   ED    ED (Normalized)    B-Cubed FS    BLEU
----------------------  -----  -----------------  ------------  ------
AntangWesternSandong    0.133              0.031         0.955   0.951
BanliangYangAn          0.371              0.091         0.901   0.854
DujiangEasternSandong   0.067              0.015         0.982   0.973
JiaoliPandong           0.438              0.100         0.863   0.847
JiarongSouthernSandong  0.152              0.035         0.945   0.944
JiuqianSouthernSandong  0.133              0.032         0.949   0.951
Pandong                 0.400              0.093         0.873   0.859
RenliEasternSandong     0.105              0.021         0.967   0.969
SanjiangEasternSandong  0.095              0.023         0.976   0.960
ShuigenCentralSandong   0.057              0.014         0.979   0.977
ShuiweiSouthernSandong  0.105              0.025         0.962   0.962
ShuiyaoSouthernSandong  0.238              0.057         0.919   0.912
TangnianYangAn          0.286              0.068         0.911   0.886
TangzhouWesternSandong  0.114              0.025         0.964   0.960
TingpaiWesternSandong   0.133              0.032         0.954   0.955
ZhongheCentralSandong   0.048              0.013         0.980   0.978
TOTAL                   0.180              0.042         0.942   0.934
```

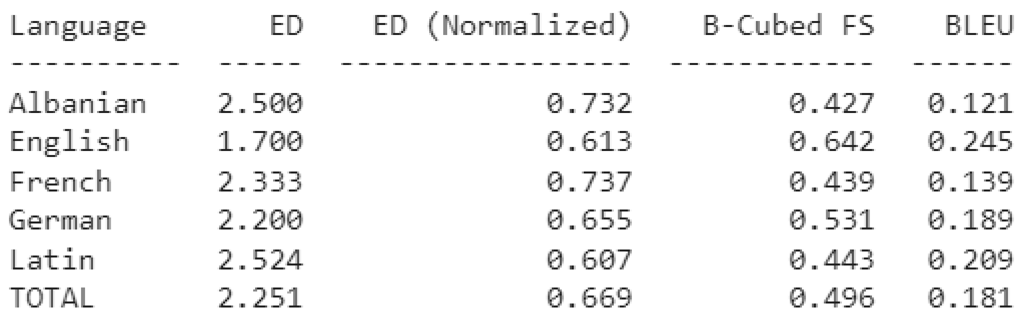
SURPRISE DATA SET

1. Dataset: Wangbai



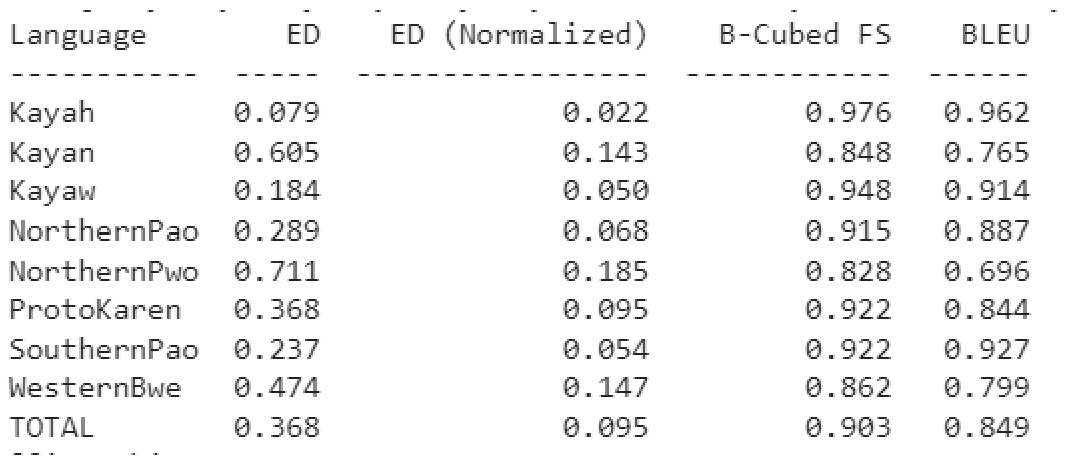
```
Language       ED    ED (Normalized)    B-Cubed FS    BLEU
----------  -----  -----------------  ------------  ------
Dashi       0.682              0.202         0.780   0.703
Ega         0.424              0.134         0.844   0.795
Enqi        0.470              0.134         0.835   0.803
Gongxing    0.636              0.185         0.781   0.716
Jinman      0.636              0.198         0.778   0.703
Jinxing     0.667              0.190         0.771   0.710
Mazhelong   0.742              0.217         0.764   0.682
ProtoBai    0.697              0.177         0.770   0.726
Tuoluo      0.500              0.134         0.837   0.792
Zhoucheng   0.409              0.129         0.834   0.809
TOTAL       0.586              0.170         0.799   0.744
[['Dashi',
```

3. Dataset: Kesslersignificance



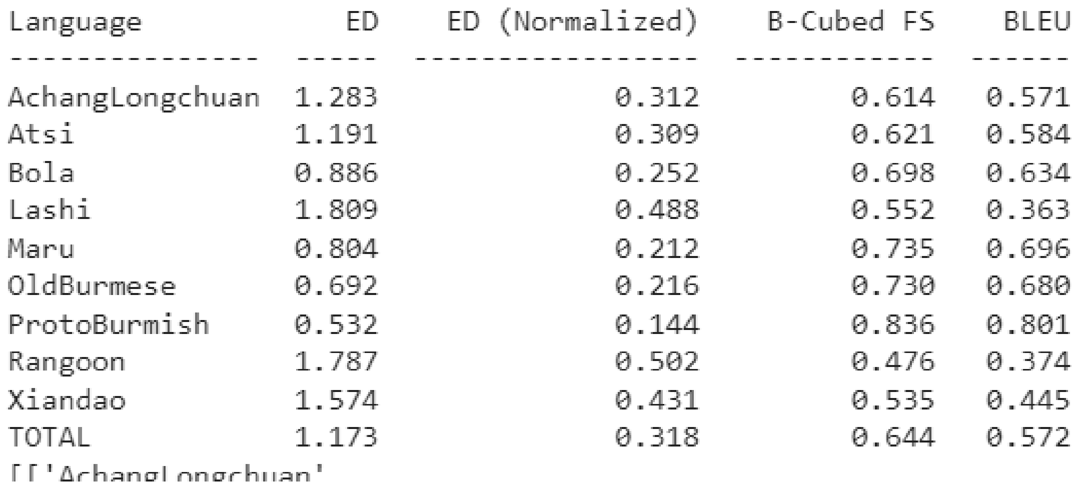
```
Language       ED    ED (Normalized)    B-Cubed FS    BLEU
----------  -----  -----------------  ------------  ------
Albanian    2.500              0.732         0.427   0.121
English     1.700              0.613         0.642   0.245
French      2.333              0.737         0.439   0.139
German      2.200              0.655         0.531   0.189
Latin       2.524              0.607         0.443   0.209
TOTAL       2.251              0.669         0.496   0.181
```

2. Dataset: Luangthongkumkaren



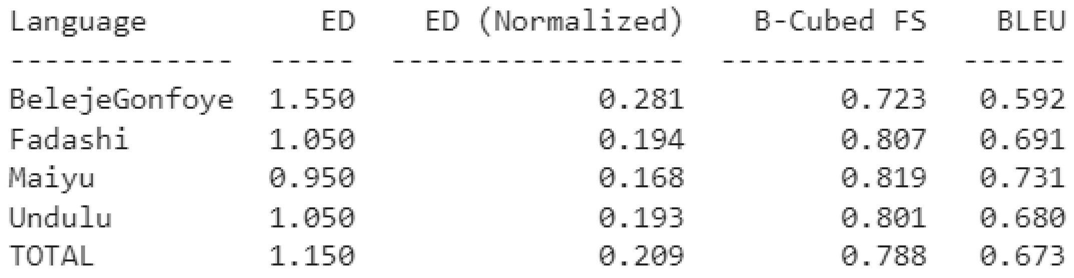
```
Language        ED    ED (Normalized)    B-Cubed FS    BLEU
-----------  -----  -----------------  ------------  ------
Kayah        0.079              0.022         0.976   0.962
Kayan        0.605              0.143         0.848   0.765
Kayaw        0.184              0.050         0.948   0.914
NorthernPao  0.289              0.068         0.915   0.887
NorthernPwo  0.711              0.185         0.828   0.696
ProtoKaren   0.368              0.095         0.922   0.844
SouthernPao  0.237              0.054         0.922   0.927
WesternBwe   0.474              0.147         0.862   0.799
TOTAL        0.368              0.095         0.903   0.849
```

4. Dataset: Hillburmish



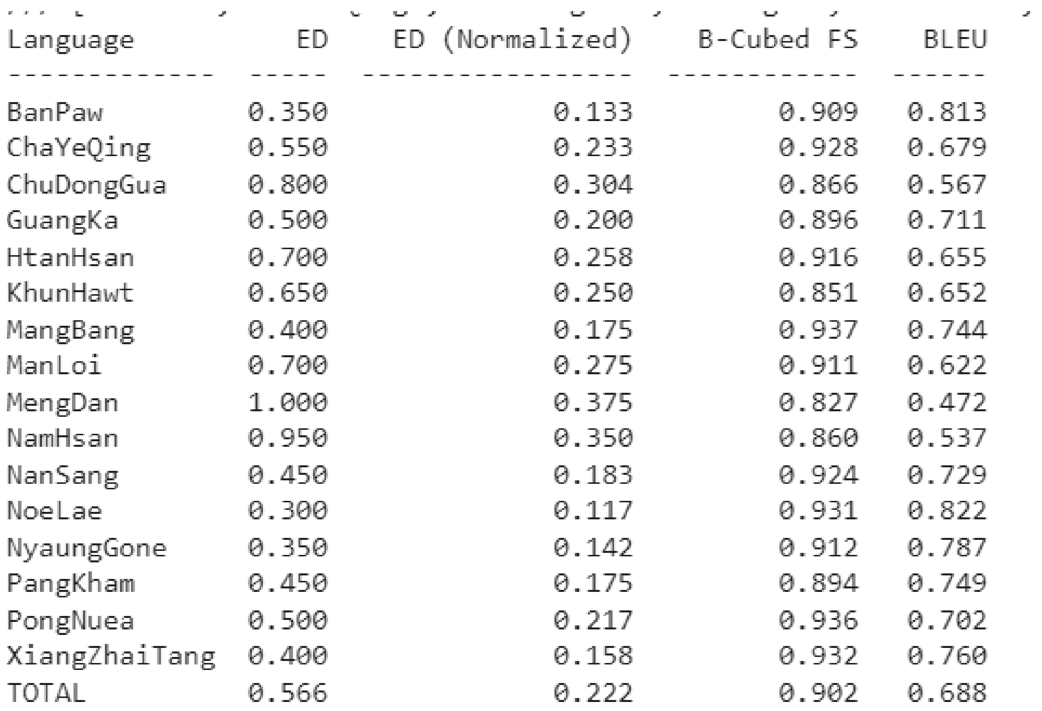
```
Language            ED    ED (Normalized)    B-Cubed FS    BLEU
---------------  -----  -----------------  ------------  ------
AchangLongchuan  1.283              0.312         0.614   0.571
Atsi             1.191              0.309         0.621   0.584
Bola             0.886              0.252         0.698   0.634
Lashi            1.809              0.488         0.552   0.363
Maru             0.804              0.212         0.735   0.696
OldBurmese       0.692              0.216         0.730   0.680
ProtoBurmish     0.532              0.144         0.836   0.801
Rangoon          1.787              0.502         0.476   0.374
Xiandao          1.574              0.431         0.535   0.445
TOTAL            1.173              0.318         0.644   0.572
[['AchangLongchuan'
```

5. Dataset: bremerberta



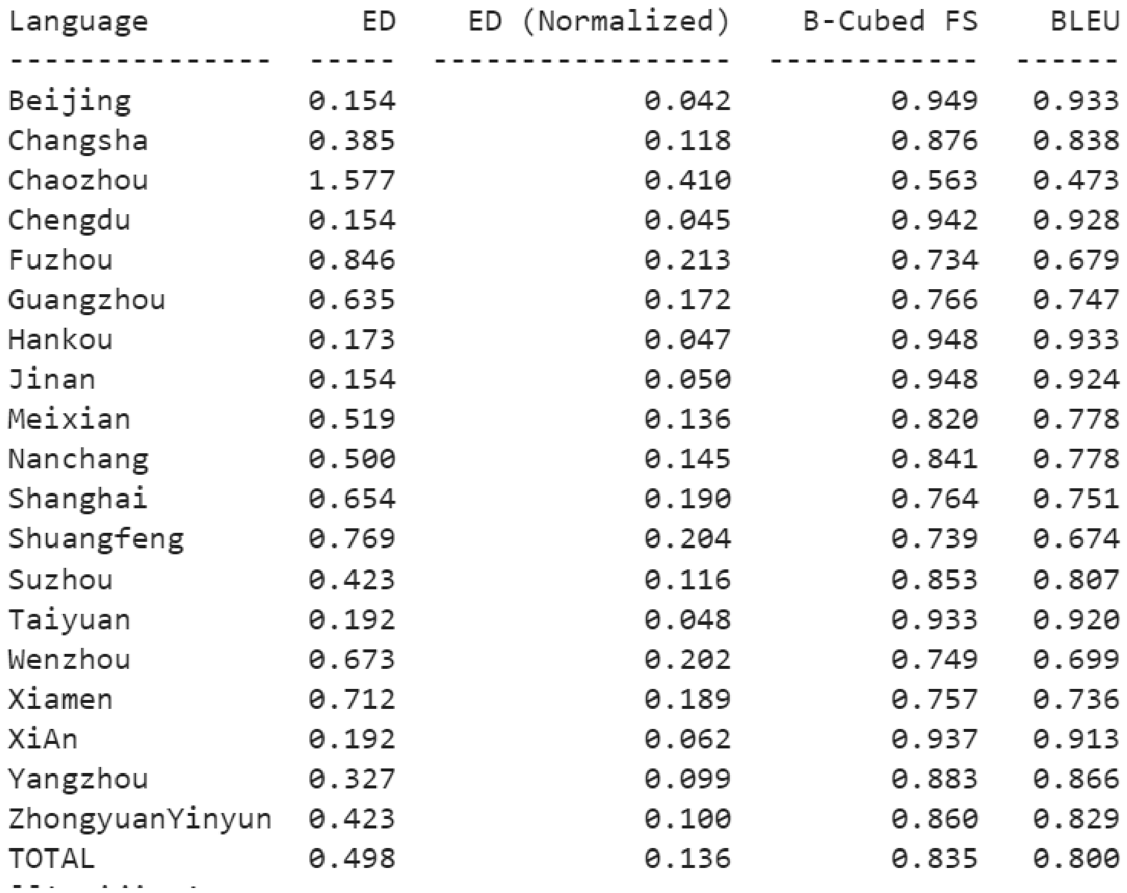
```
Language          ED    ED (Normalized)    B-Cubed FS    BLEU
-------------  -----  -----------------  ------------  ------
BelejeGonfoye  1.550              0.281         0.723   0.592
Fadashi        1.050              0.194         0.807   0.691
Maiyu          0.950              0.168         0.819   0.731
Undulu         1.050              0.193         0.801   0.680
TOTAL          1.150              0.209         0.788   0.673
```

6. Dataset: deepadungpalaung



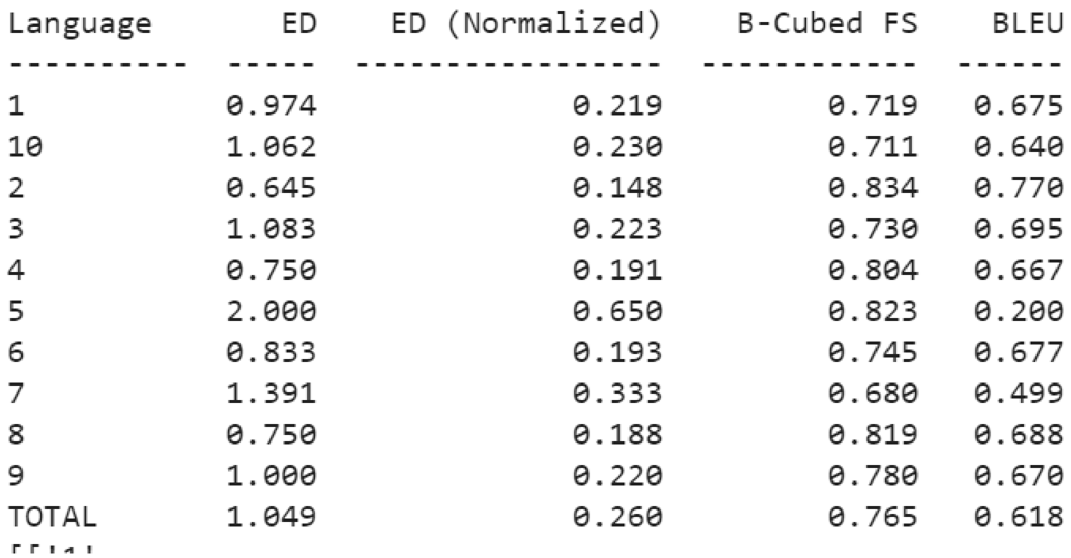
```
Language          ED    ED (Normalized)    B-Cubed FS    BLEU
-------------  -----  -----------------  ------------  ------
BanPaw         0.350              0.133         0.909   0.813
ChaYeQing      0.550              0.233         0.928   0.679
ChuDongGua     0.800              0.304         0.866   0.567
GuangKa        0.500              0.200         0.896   0.711
HtanHsan       0.700              0.258         0.916   0.655
KhunHawt       0.650              0.250         0.851   0.652
MangBang       0.400              0.175         0.937   0.744
ManLoi         0.700              0.275         0.911   0.622
MengDan        1.000              0.375         0.827   0.472
NamHsan        0.950              0.350         0.860   0.537
NanSang        0.450              0.183         0.924   0.729
NoeLae         0.300              0.117         0.931   0.822
NyaungGone     0.350              0.142         0.912   0.787
PangKham       0.450              0.175         0.894   0.749
PongNuea       0.500              0.217         0.936   0.702
XiangZhaiTang  0.400              0.158         0.932   0.760
TOTAL          0.566              0.222         0.902   0.688
```

7. Dataset: beidazihui



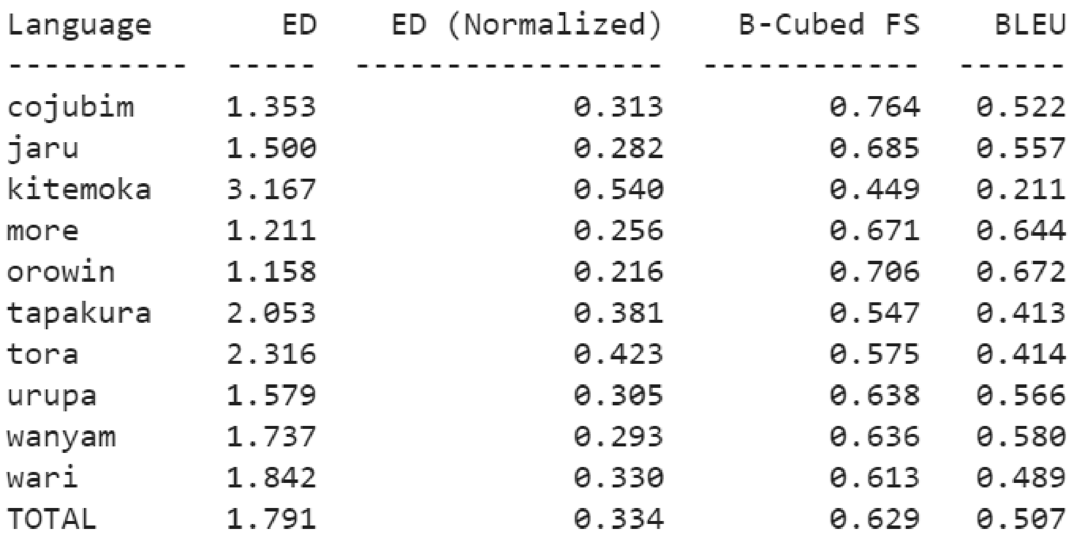
```
Language             ED    ED (Normalized)    B-Cubed FS    BLEU
---------------  -----  -----------------  ------------  ------
Beijing          0.154              0.042         0.949   0.933
Changsha         0.385              0.118         0.876   0.838
Chaozhou         1.577              0.410         0.563   0.473
Chengdu          0.154              0.045         0.942   0.928
Fuzhou           0.846              0.213         0.734   0.679
Guangzhou        0.635              0.172         0.766   0.747
Hankou           0.173              0.047         0.948   0.933
Jinan            0.154              0.050         0.948   0.924
Meixian          0.519              0.136         0.820   0.778
Nanchang         0.500              0.145         0.841   0.778
Shanghai         0.654              0.190         0.764   0.751
Shuangfeng       0.769              0.204         0.739   0.674
Suzhou           0.423              0.116         0.853   0.807
Taiyuan          0.192              0.048         0.933   0.920
Wenzhou          0.673              0.202         0.749   0.699
Xiamen           0.712              0.189         0.757   0.736
XiAn             0.192              0.062         0.937   0.913
Yangzhou         0.327              0.099         0.883   0.866
ZhongyuanYinyun  0.423              0.100         0.860   0.829
TOTAL            0.498              0.136         0.835   0.800
```

8. Dataset: bantubvd



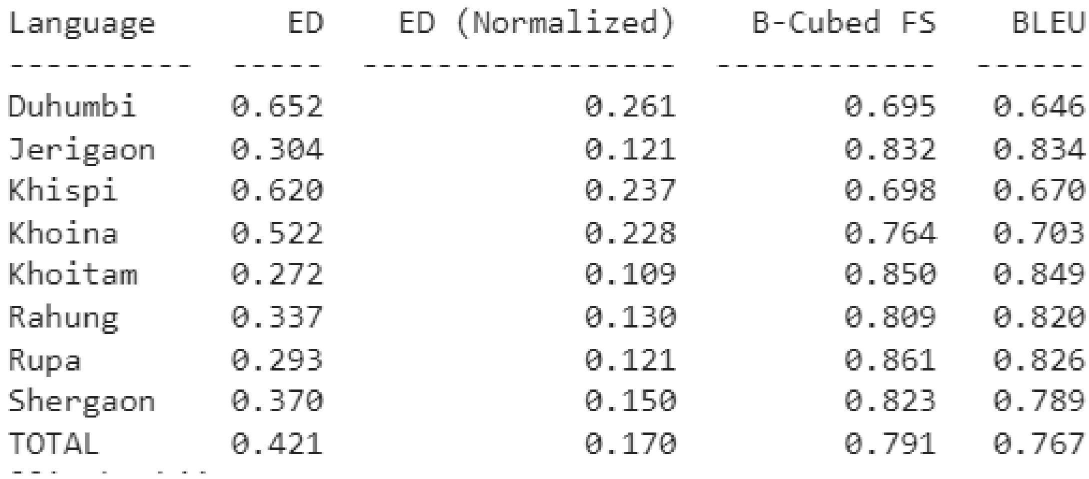
```
Language       ED    ED (Normalized)    B-Cubed FS    BLEU
----------  -----  -----------------  ------------  ------
1           0.974              0.219         0.719   0.675
10          1.062              0.230         0.711   0.640
2           0.645              0.148         0.834   0.770
3           1.083              0.223         0.730   0.695
4           0.750              0.191         0.804   0.667
5           2.000              0.650         0.823   0.200
6           0.833              0.193         0.745   0.677
7           1.391              0.333         0.680   0.499
8           0.750              0.188         0.819   0.688
9           1.000              0.220         0.780   0.670
TOTAL       1.049              0.260         0.765   0.618
```

9. Dataset: brichallchapacuran



```
Language       ED    ED (Normalized)    B-Cubed FS    BLEU
----------  -----  -----------------  ------------  ------
cojubim     1.353              0.313         0.764   0.522
jaru        1.500              0.282         0.685   0.557
kitemoka    3.167              0.540         0.449   0.211
more        1.211              0.256         0.671   0.644
orowin      1.158              0.216         0.706   0.672
tapakura    2.053              0.381         0.547   0.413
tora        2.316              0.423         0.575   0.414
urupa       1.579              0.305         0.638   0.566
wanyam      1.737              0.293         0.636   0.580
wari        1.842              0.330         0.613   0.489
TOTAL       1.791              0.334         0.629   0.507
```

10. Dataset: bodtkhobwa

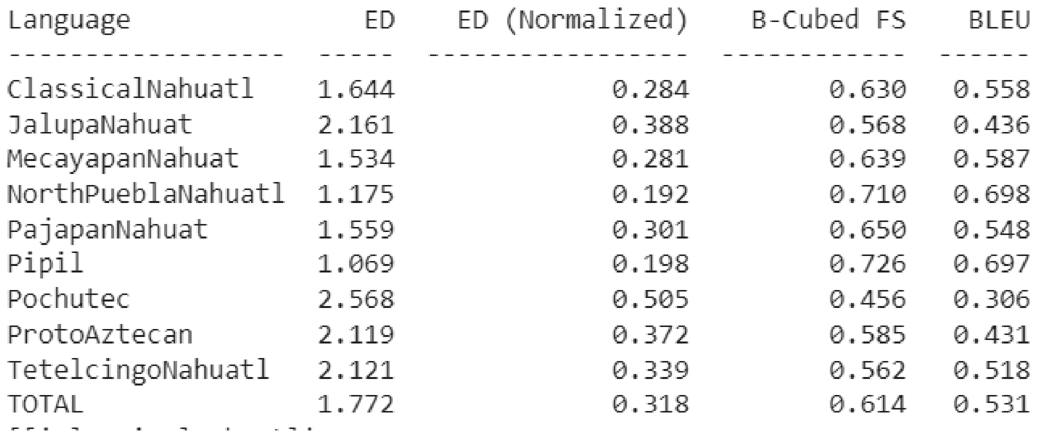


```
Language       ED    ED (Normalized)    B-Cubed FS    BLEU
----------  -----  -----------------  ------------  ------
Duhumbi     0.652              0.261         0.695   0.646
Jerigaon    0.304              0.121         0.832   0.834
Khispi      0.620              0.237         0.698   0.670
Khoina      0.522              0.228         0.764   0.703
Khoitam     0.272              0.109         0.850   0.849
Rahung      0.337              0.130         0.809   0.820
Rupa        0.293              0.121         0.861   0.826
Shergaon    0.370              0.150         0.823   0.789
TOTAL       0.421              0.170         0.791   0.767
```

Results for the split - 0.50

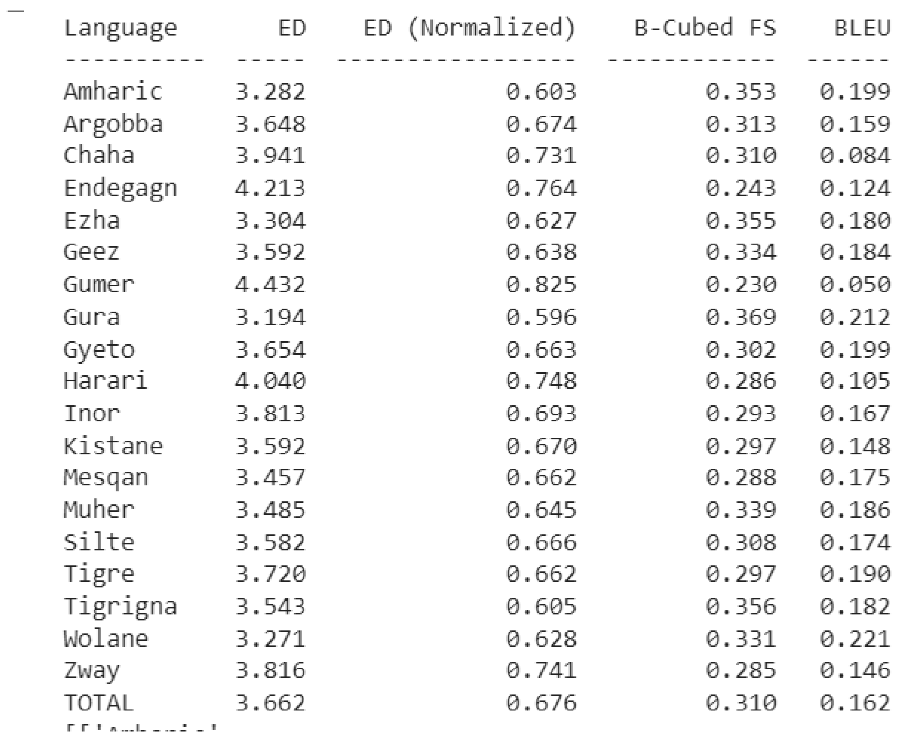
TRAINING DATA SET

1. Dataset: davletshinaztecan



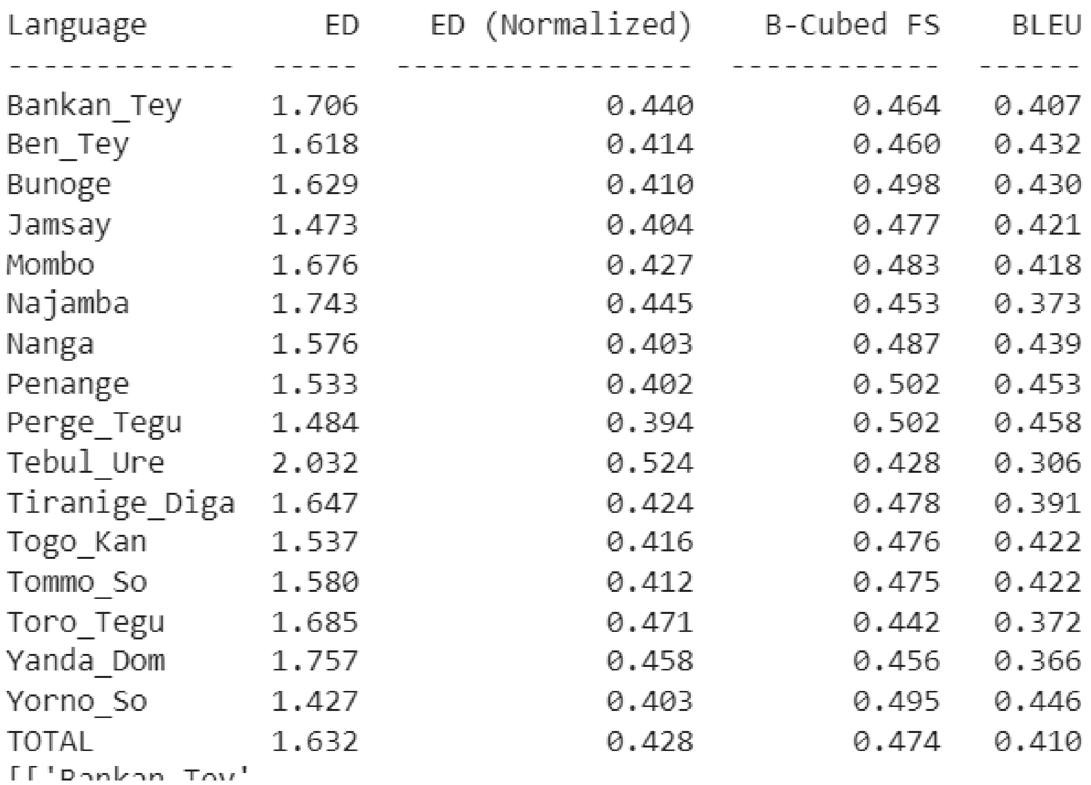
```
Language              ED    ED (Normalized)    B-Cubed FS    BLEU
------------------  -----  -----------------  ------------  ------
ClassicalNahuatl    1.644              0.284         0.630   0.558
JalupaNahuat        2.161              0.388         0.568   0.436
MecayapanNahuat     1.534              0.281         0.639   0.587
NorthPueblaNahuatl  1.175              0.192         0.710   0.698
PajapanNahuat       1.559              0.301         0.650   0.548
Pipil               1.069              0.198         0.726   0.697
Pochutec            2.568              0.505         0.456   0.306
ProtoAztecan        2.119              0.372         0.585   0.431
TetelcingoNahuatl   2.121              0.339         0.562   0.518
TOTAL               1.772              0.318         0.614   0.531
```

2.Dataset : felekesemitic



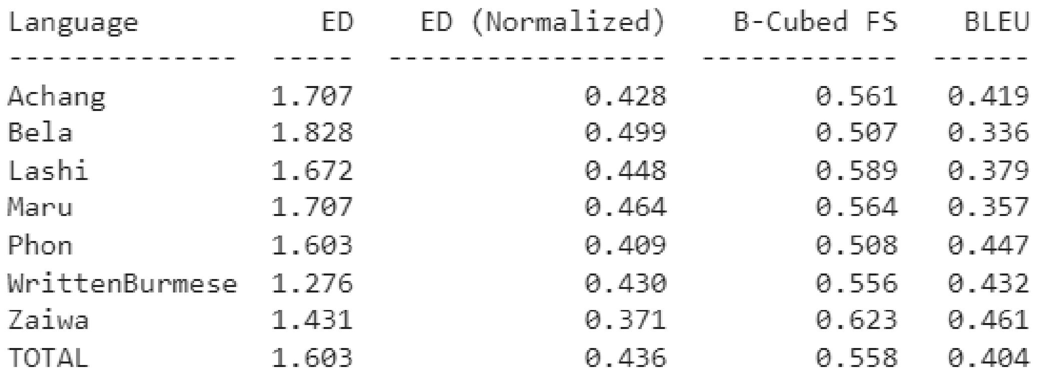
```
Language      ED    ED (Normalized)    B-Cubed FS    BLEU
----------  -----  -----------------  ------------  ------
Amharic     3.282              0.603         0.353   0.199
Argobba     3.648              0.674         0.313   0.159
Chaha       3.941              0.731         0.310   0.084
Endegagn    4.213              0.764         0.243   0.124
Ezha        3.304              0.627         0.355   0.180
Geez        3.592              0.638         0.334   0.184
Gumer       4.432              0.825         0.230   0.050
Gura        3.194              0.596         0.369   0.212
Gyeto       3.654              0.663         0.302   0.199
Harari      4.040              0.748         0.286   0.105
Inor        3.813              0.693         0.293   0.167
Kistane     3.592              0.670         0.297   0.148
Mesqan      3.457              0.662         0.288   0.175
Muher       3.485              0.645         0.339   0.186
Silte       3.582              0.666         0.308   0.174
Tigre       3.720              0.662         0.297   0.190
Tigrigna    3.543              0.605         0.356   0.182
Wolane      3.271              0.628         0.331   0.221
Zway        3.816              0.741         0.285   0.146
TOTAL       3.662              0.676         0.310   0.162
[['Amharic'
```

3.Dataset: hantganbangime



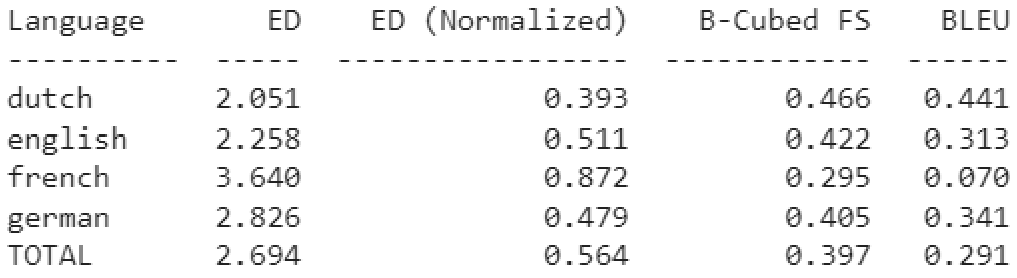
```
Language            ED    ED (Normalized)    B-Cubed FS    BLEU
-------------  -----  -----------------  ------------  ------
Bankan_Tey     1.706              0.440         0.464   0.407
Ben_Tey        1.618              0.414         0.460   0.432
Bunoge         1.629              0.410         0.498   0.430
Jamsay         1.473              0.404         0.477   0.421
Mombo          1.676              0.427         0.483   0.418
Najamba        1.743              0.445         0.453   0.373
Nanga          1.576              0.403         0.487   0.439
Penange        1.533              0.402         0.502   0.453
Perge_Tegu     1.484              0.394         0.502   0.458
Tebul_Ure      2.032              0.524         0.428   0.306
Tiranige_Diga  1.647              0.424         0.478   0.391
Togo_Kan       1.537              0.416         0.476   0.422
Tommo_So       1.580              0.412         0.475   0.422
Toro_Tegu      1.685              0.471         0.442   0.372
Yanda_Dom      1.757              0.458         0.456   0.366
Yorno_So       1.427              0.403         0.495   0.446
TOTAL          1.632              0.428         0.474   0.410
[['Bankan_Tey'
```

4. Dataset: mannburmish



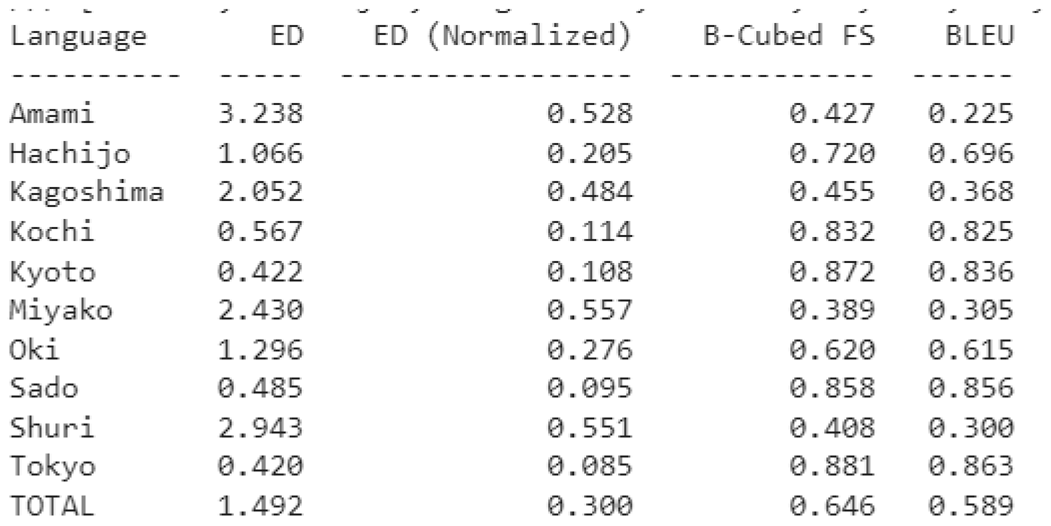
```
Language             ED    ED (Normalized)    B-Cubed FS    BLEU
--------------  -----  -----------------  ------------  ------
Achang          1.707              0.428         0.561   0.419
Bela            1.828              0.499         0.507   0.336
Lashi           1.672              0.448         0.589   0.379
Maru            1.707              0.464         0.564   0.357
Phon            1.603              0.409         0.508   0.447
WrittenBurmese  1.276              0.430         0.556   0.432
Zaiwa           1.431              0.371         0.623   0.461
TOTAL           1.603              0.436         0.558   0.404
```

5. Dataset: listsamplesize



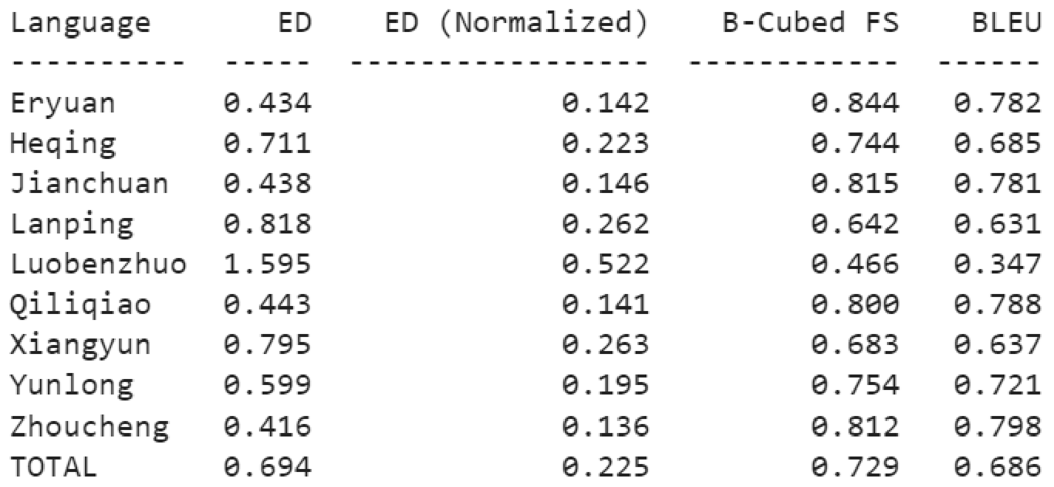
```
Language         ED    ED (Normalized)    B-Cubed FS    BLEU
----------  -----  -----------------  ------------  ------
dutch       2.051              0.393         0.466   0.441
english     2.258              0.511         0.422   0.313
french      3.640              0.872         0.295   0.070
german      2.826              0.479         0.405   0.341
TOTAL       2.694              0.564         0.397   0.291
```

6. Dataset: hattorijaponic



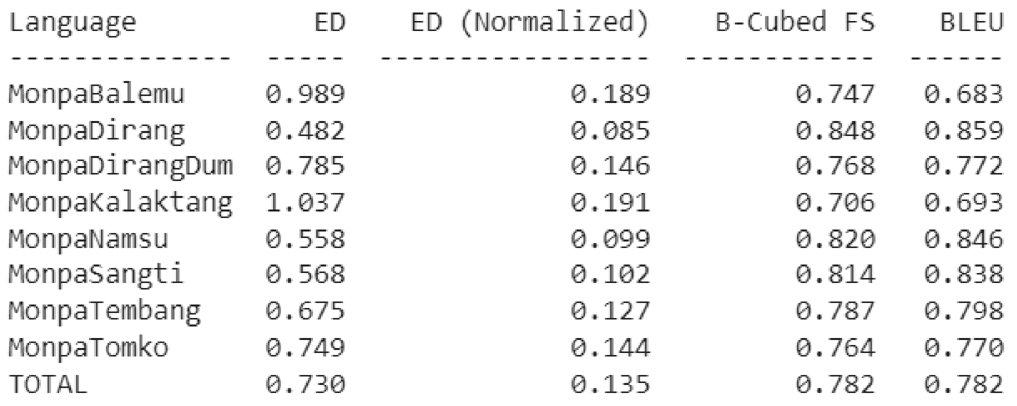
```
Language          ED    ED (Normalized)    B-Cubed FS    BLEU
----------     -----  -----------------  ------------  ------
Amami          3.238              0.528         0.427   0.225
Hachijo        1.066              0.205         0.720   0.696
Kagoshima      2.052              0.484         0.455   0.368
Kochi          0.567              0.114         0.832   0.825
Kyoto          0.422              0.108         0.872   0.836
Miyako         2.430              0.557         0.389   0.305
Oki            1.296              0.276         0.620   0.615
Sado           0.485              0.095         0.858   0.856
Shuri          2.943              0.551         0.408   0.300
Tokyo          0.420              0.085         0.881   0.863
TOTAL          1.492              0.300         0.646   0.589
```

7.Language: allenbai



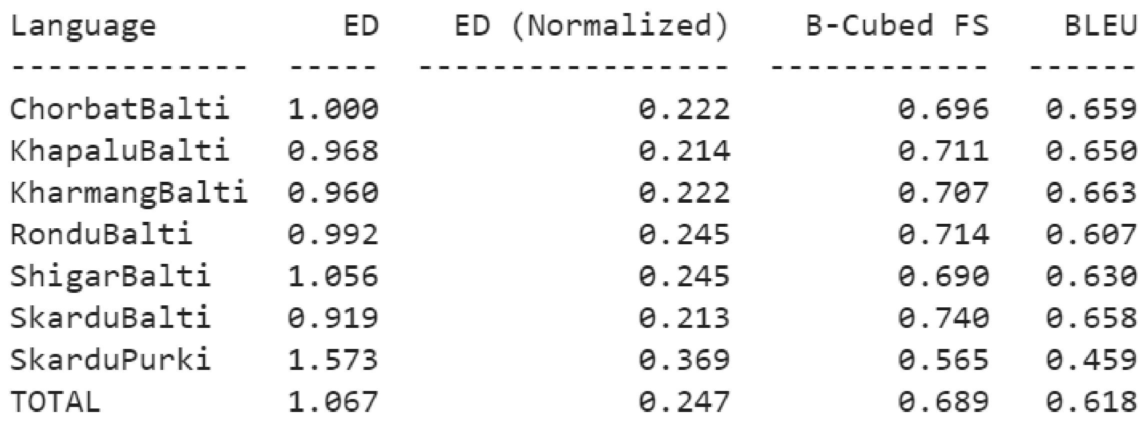
```
Language          ED    ED (Normalized)    B-Cubed FS    BLEU
----------     -----  -----------------  ------------  ------
Eryuan         0.434              0.142         0.844   0.782
Heqing         0.711              0.223         0.744   0.685
Jianchuan      0.438              0.146         0.815   0.781
Lanping        0.818              0.262         0.642   0.631
Luobenzhuo     1.595              0.522         0.466   0.347
Qiliqiao       0.443              0.141         0.800   0.788
Xiangyun       0.795              0.263         0.683   0.637
Yunlong        0.599              0.195         0.754   0.721
Zhoucheng      0.416              0.136         0.812   0.798
TOTAL          0.694              0.225         0.729   0.686
```

8.Language: abrahammonpa



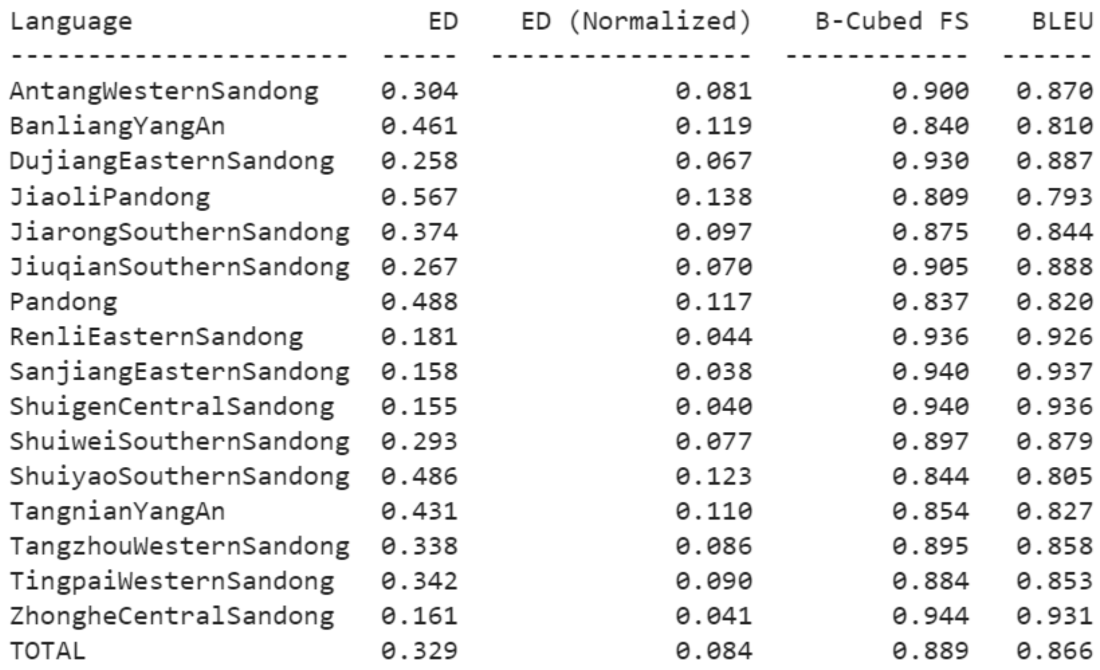
```
Language             ED    ED (Normalized)    B-Cubed FS    BLEU
-------------     -----  -----------------  ------------  ------
MonpaBalemu       0.989              0.189         0.747   0.683
MonpaDirang       0.482              0.085         0.848   0.859
MonpaDirangDum    0.785              0.146         0.768   0.772
MonpaKalaktang    1.037              0.191         0.706   0.693
MonpaNamsu        0.558              0.099         0.820   0.846
MonpaSangti       0.568              0.102         0.814   0.838
MonpaTembang      0.675              0.127         0.787   0.798
MonpaTomko        0.749              0.144         0.764   0.770
TOTAL             0.730              0.135         0.782   0.782
```

9.Language: backstromnorthernpakistan



```
Language          ED    ED (Normalized)    B-Cubed FS    BLEU
-------------  -----  -----------------  ------------  ------
ChorbatBalti   1.000              0.222         0.696   0.659
KhapaluBalti   0.968              0.214         0.711   0.650
KharmangBalti  0.960              0.222         0.707   0.663
RonduBalti     0.992              0.245         0.714   0.607
ShigarBalti    1.056              0.245         0.690   0.630
SkarduBalti    0.919              0.213         0.740   0.658
SkarduPurki    1.573              0.369         0.565   0.459
TOTAL          1.067              0.247         0.689   0.618
```

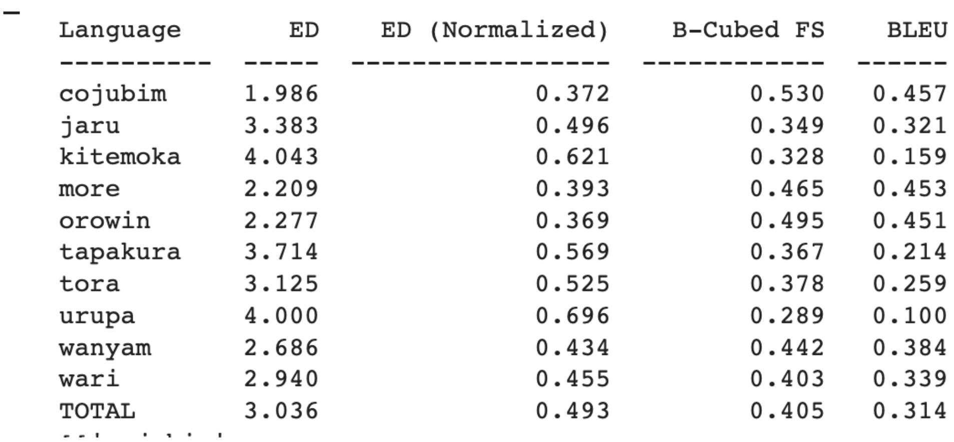
10.Language: castrosui



```
Language                   ED    ED (Normalized)    B-Cubed FS    BLEU
----------------------  -----  -----------------  ------------  ------
AntangWesternSandong    0.304              0.081         0.900   0.870
BanliangYangAn          0.461              0.119         0.840   0.810
DujiangEasternSandong   0.258              0.067         0.930   0.887
JiaoliPandong           0.567              0.138         0.809   0.793
JiarongSouthernSandong  0.374              0.097         0.875   0.844
JiuqianSouthernSandong  0.267              0.070         0.905   0.888
Pandong                 0.488              0.117         0.837   0.820
RenliEasternSandong     0.181              0.044         0.936   0.926
SanjiangEasternSandong  0.158              0.038         0.940   0.937
ShuigenCentralSandong   0.155              0.040         0.940   0.936
ShuiweiSouthernSandong  0.293              0.077         0.897   0.879
ShuiyaoSouthernSandong  0.486              0.123         0.844   0.805
TangnianYangAn          0.431              0.110         0.854   0.827
TangzhouWesternSandong  0.338              0.086         0.895   0.858
TingpaiWesternSandong   0.342              0.090         0.884   0.853
ZhongheCentralSandong   0.161              0.041         0.944   0.931
TOTAL                   0.329              0.084         0.889   0.866
```

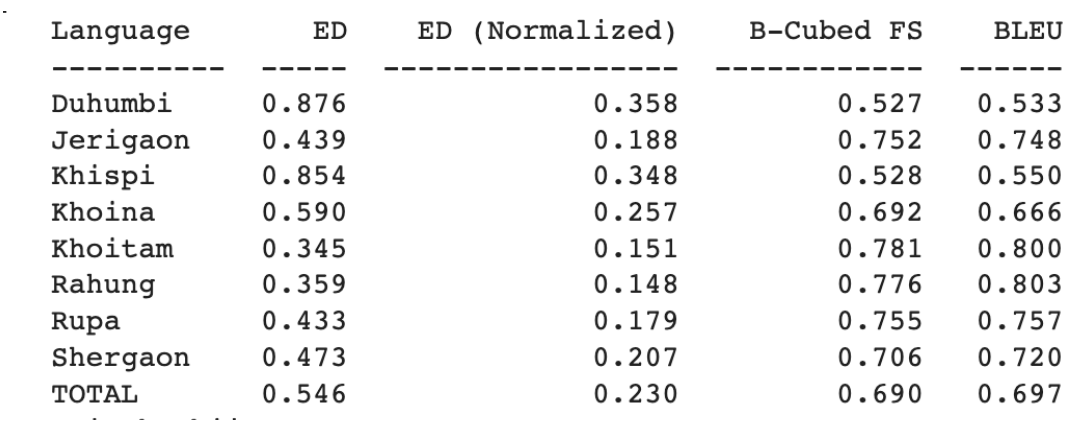
SURPRISE DATA:0.50 Proportion

1.Dataset: birchallchapacuran



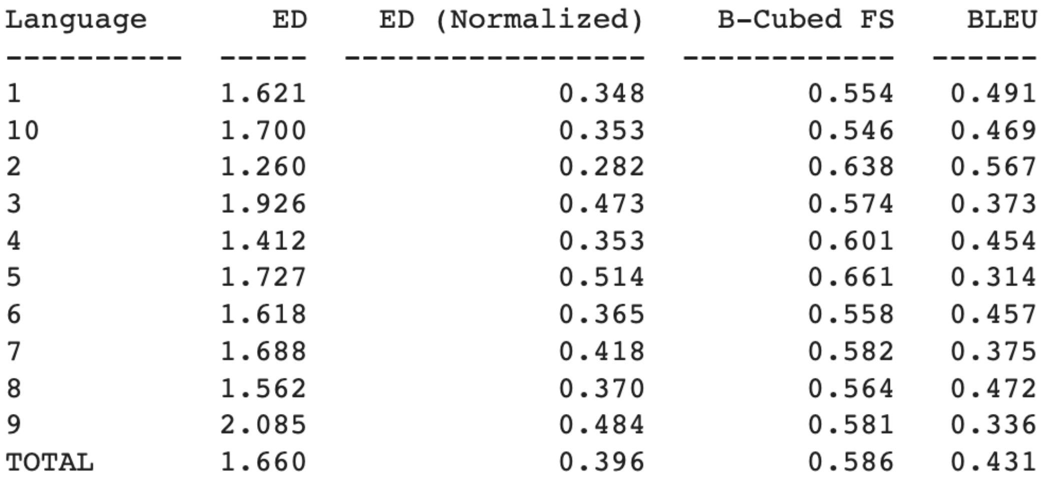
```
Language       ED    ED (Normalized)    B-Cubed FS    BLEU
----------  -----  -----------------  ------------  ------
cojubim     1.986              0.372         0.530   0.457
jaru        3.383              0.496         0.349   0.321
kitemoka    4.043              0.621         0.328   0.159
more        2.209              0.393         0.465   0.453
orowin      2.277              0.369         0.495   0.451
tapakura    3.714              0.569         0.367   0.214
tora        3.125              0.525         0.378   0.259
urupa       4.000              0.696         0.289   0.100
wanyam      2.686              0.434         0.442   0.384
wari        2.940              0.455         0.403   0.339
TOTAL       3.036              0.493         0.405   0.314
```

2.Dataset : bodtkhobwa



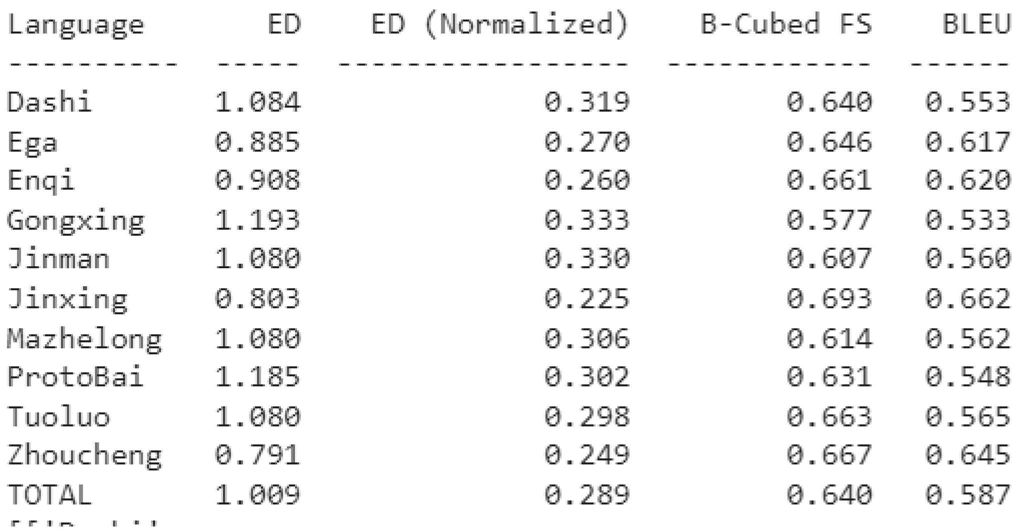
```
Language       ED    ED (Normalized)    B-Cubed FS    BLEU
----------  -----  -----------------  ------------  ------
Duhumbi     0.876              0.358         0.527   0.533
Jerigaon    0.439              0.188         0.752   0.748
Khispi      0.854              0.348         0.528   0.550
Khoina      0.590              0.257         0.692   0.666
Khoitam     0.345              0.151         0.781   0.800
Rahung      0.359              0.148         0.776   0.803
Rupa        0.433              0.179         0.755   0.757
Shergaon    0.473              0.207         0.706   0.720
TOTAL       0.546              0.230         0.690   0.697
```

3.Dataset : bantubvd



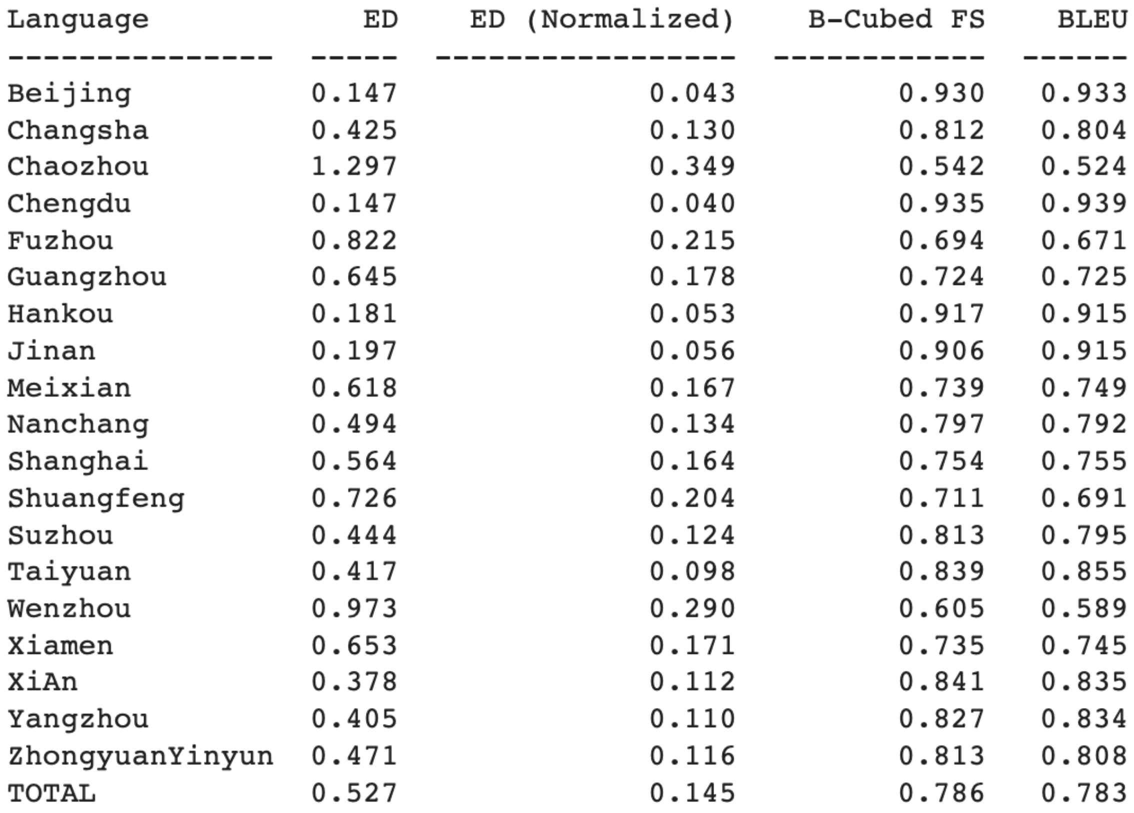
```
Language       ED    ED (Normalized)    B-Cubed FS    BLEU
----------  -----  -----------------  ------------  ------
1           1.621              0.348         0.554   0.491
10          1.700              0.353         0.546   0.469
2           1.260              0.282         0.638   0.567
3           1.926              0.473         0.574   0.373
4           1.412              0.353         0.601   0.454
5           1.727              0.514         0.661   0.314
6           1.618              0.365         0.558   0.457
7           1.688              0.418         0.582   0.375
8           1.562              0.370         0.564   0.472
9           2.085              0.484         0.581   0.336
TOTAL       1.660              0.396         0.586   0.431
```

4. Dataset: Wangbai



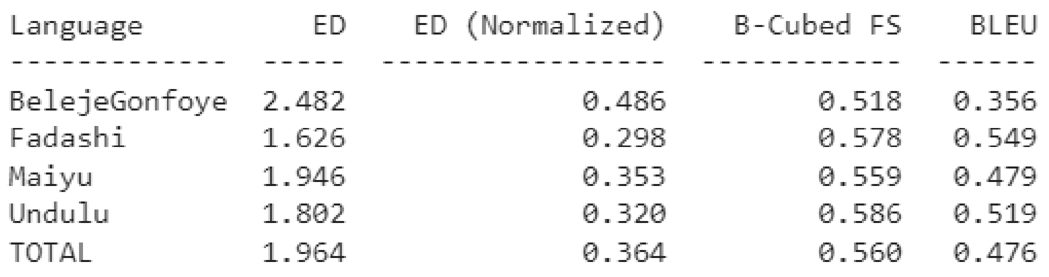
```
Language       ED    ED (Normalized)    B-Cubed FS    BLEU
---------    -----  -----------------  ------------  ------
Dashi        1.084              0.319         0.640   0.553
Ega          0.885              0.270         0.646   0.617
Enqi         0.908              0.260         0.661   0.620
Gongxing     1.193              0.333         0.577   0.533
Jinman       1.080              0.330         0.607   0.560
Jinxing      0.803              0.225         0.693   0.662
Mazhelong    1.080              0.306         0.614   0.562
ProtoBai     1.185              0.302         0.631   0.548
Tuoluo       1.080              0.298         0.663   0.565
Zhoucheng    0.791              0.249         0.667   0.645
TOTAL        1.009              0.289         0.640   0.587
```

5. Dataset: beidazihui



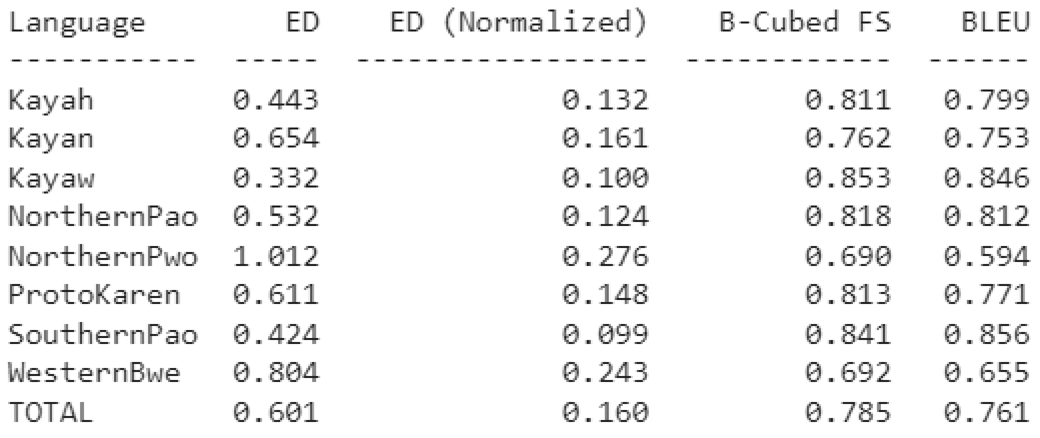
```
Language             ED    ED (Normalized)    B-Cubed FS    BLEU
---------------    -----  -----------------  ------------  ------
Beijing            0.147              0.043         0.930   0.933
Changsha           0.425              0.130         0.812   0.804
Chaozhou           1.297              0.349         0.542   0.524
Chengdu            0.147              0.040         0.935   0.939
Fuzhou             0.822              0.215         0.694   0.671
Guangzhou          0.645              0.178         0.724   0.725
Hankou             0.181              0.053         0.917   0.915
Jinan              0.197              0.056         0.906   0.915
Meixian            0.618              0.167         0.739   0.749
Nanchang           0.494              0.134         0.797   0.792
Shanghai           0.564              0.164         0.754   0.755
Shuangfeng         0.726              0.204         0.711   0.691
Suzhou             0.444              0.124         0.813   0.795
Taiyuan            0.417              0.098         0.839   0.855
Wenzhou            0.973              0.290         0.605   0.589
Xiamen             0.653              0.171         0.735   0.745
XiAn               0.378              0.112         0.841   0.835
Yangzhou           0.405              0.110         0.827   0.834
ZhongyuanYinyun    0.471              0.116         0.813   0.808
TOTAL              0.527              0.145         0.786   0.783
```

6. Dataset: bremerberta



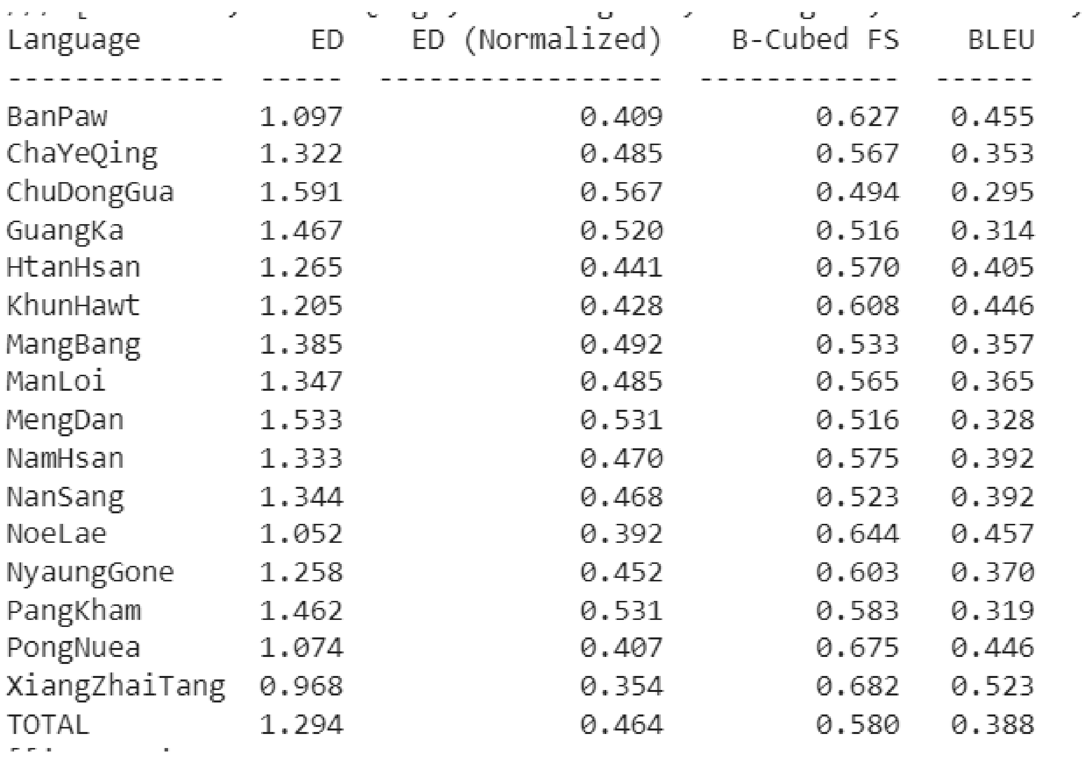
```
Language          ED    ED (Normalized)    B-Cubed FS    BLEU
-------------  -----  -----------------  ------------  ------
BelejeGonfoye  2.482              0.486         0.518   0.356
Fadashi        1.626              0.298         0.578   0.549
Maiyu          1.946              0.353         0.559   0.479
Undulu         1.802              0.320         0.586   0.519
TOTAL          1.964              0.364         0.560   0.476
```

7. Dataset: luangthongkumkaren



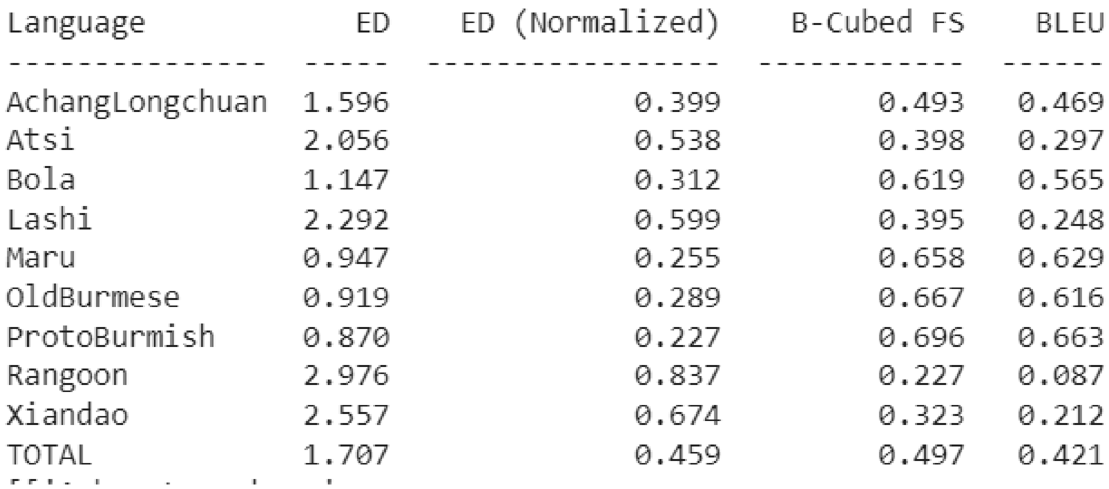
```
Language        ED    ED (Normalized)    B-Cubed FS    BLEU
-----------  -----  -----------------  ------------  ------
Kayah        0.443              0.132         0.811   0.799
Kayan        0.654              0.161         0.762   0.753
Kayaw        0.332              0.100         0.853   0.846
NorthernPao  0.532              0.124         0.818   0.812
NorthernPwo  1.012              0.276         0.690   0.594
ProtoKaren   0.611              0.148         0.813   0.771
SouthernPao  0.424              0.099         0.841   0.856
WesternBwe   0.804              0.243         0.692   0.655
TOTAL        0.601              0.160         0.785   0.761
```

8. Dataset: deepadungpalaung



```
Language         ED    ED (Normalized)    B-Cubed FS    BLEU
------------  -----  -----------------  ------------  ------
BanPaw        1.097              0.409         0.627   0.455
ChaYeQing     1.322              0.485         0.567   0.353
ChuDongGua    1.591              0.567         0.494   0.295
GuangKa       1.467              0.520         0.516   0.314
HtanHsan      1.265              0.441         0.570   0.405
KhunHawt      1.205              0.428         0.608   0.446
MangBang      1.385              0.492         0.533   0.357
ManLoi        1.347              0.485         0.565   0.365
MengDan       1.533              0.531         0.516   0.328
NamHsan       1.333              0.470         0.575   0.392
NanSang       1.344              0.468         0.523   0.392
NoeLae        1.052              0.392         0.644   0.457
NyaungGone    1.258              0.452         0.603   0.370
PangKham      1.462              0.531         0.583   0.319
PongNuea      1.074              0.407         0.675   0.446
XiangZhaiTang 0.968              0.354         0.682   0.523
TOTAL         1.294              0.464         0.580   0.388
```

9. Dataset: hillburmish

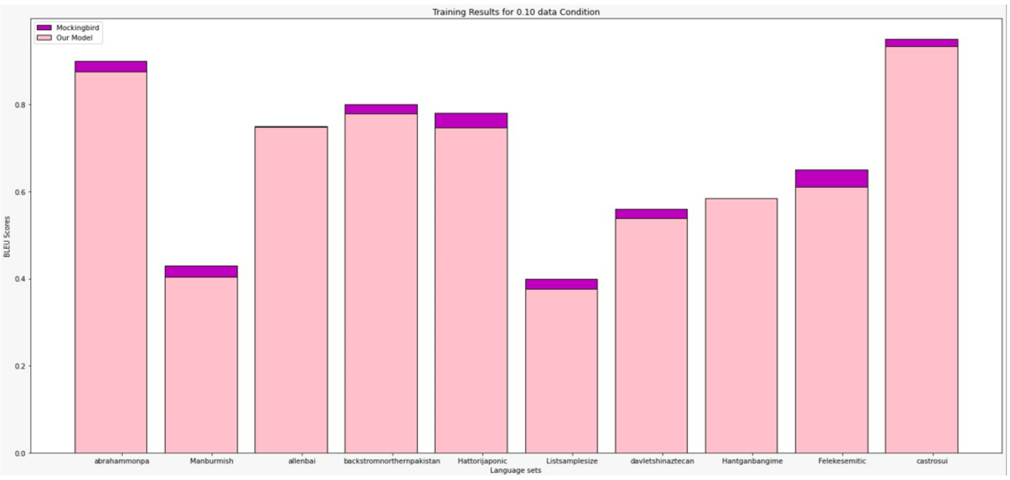


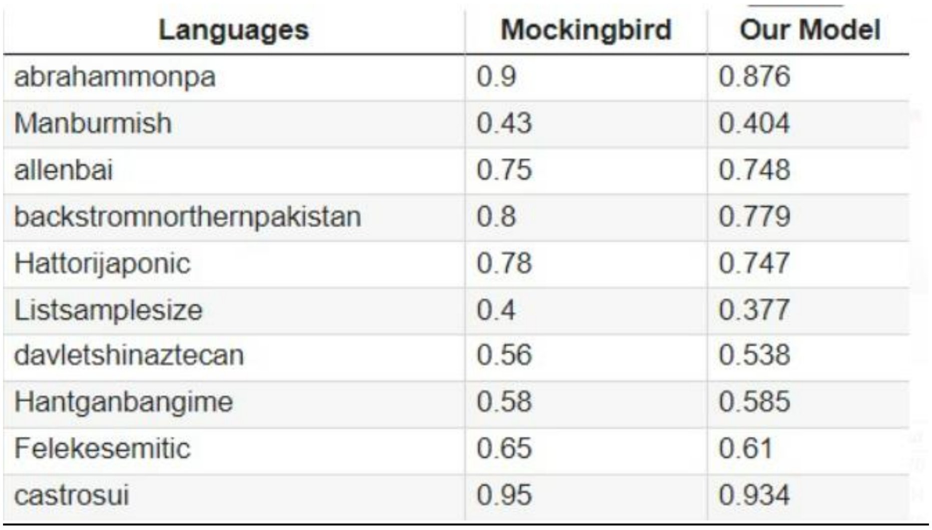
```
Language            ED    ED (Normalized)    B-Cubed FS    BLEU
---------------  -----  -----------------  ------------  ------
AchangLongchuan  1.596              0.399         0.493   0.469
Atsi             2.056              0.538         0.398   0.297
Bola             1.147              0.312         0.619   0.565
Lashi            2.292              0.599         0.395   0.248
Maru             0.947              0.255         0.658   0.629
OldBurmese       0.919              0.289         0.667   0.616
ProtoBurmish     0.870              0.227         0.696   0.663
Rangoon          2.976              0.837         0.227   0.087
Xiandao          2.557              0.674         0.323   0.212
TOTAL            1.707              0.459         0.497   0.421
```

10. Dataset : kesslersignificance

COMPARISON OF OUR RESULTS ON BASELINE:

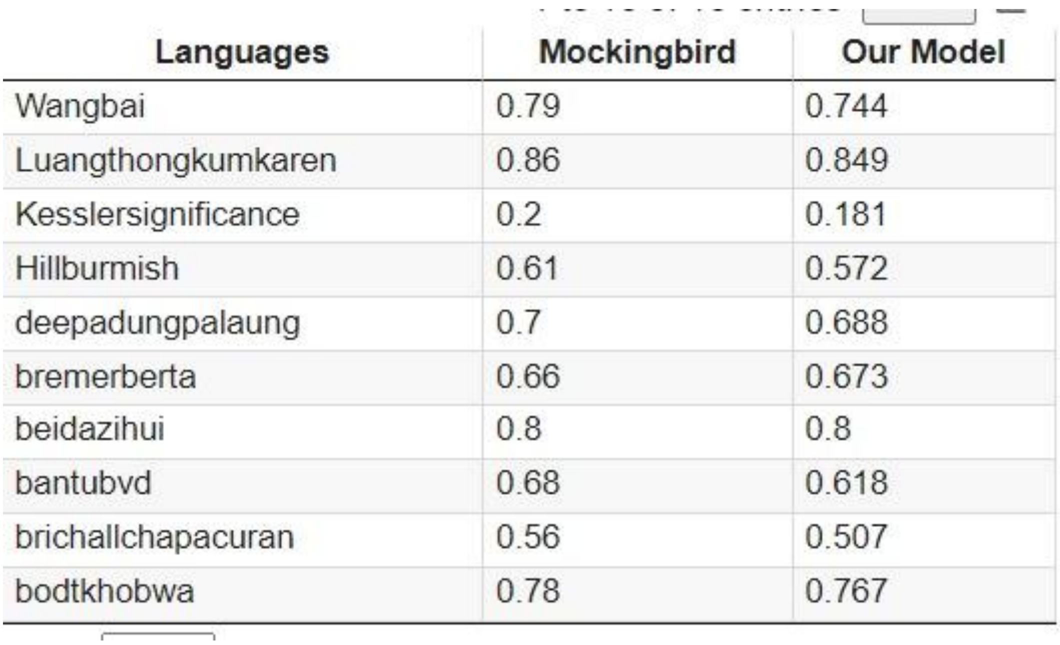
Training Data Set



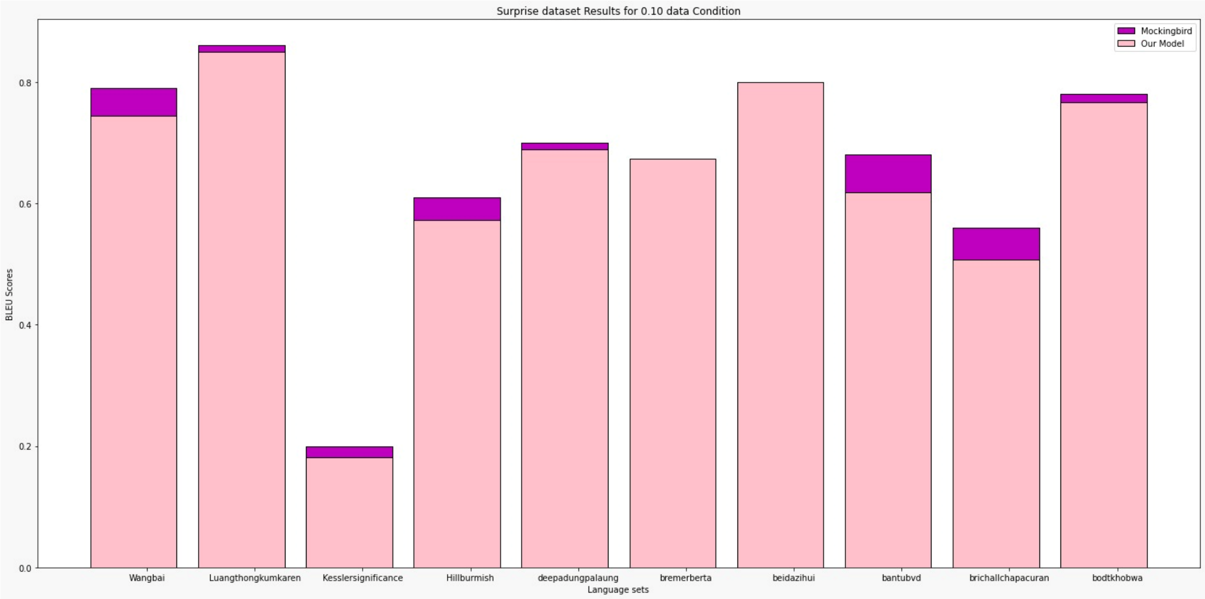


Languages	Mockingbird	Our Model
abrahammonpa	0.9	0.876
Manburmish	0.43	0.404
allenbai	0.75	0.748
backstromnorthernpakistan	0.8	0.779
Hattorijaponic	0.78	0.747
Listsamplesize	0.4	0.377
davletshinaztecan	0.56	0.538
Hantganbangime	0.58	0.585
Felekesemitic	0.65	0.61
castrosui	0.95	0.934

Surprise Data Set



Languages	Mockingbird	Our Model
Wangbai	0.79	0.744
Luangthongkumkaren	0.86	0.849
Kesslersignificance	0.2	0.181
Hillburmish	0.61	0.572
deepadungpalaung	0.7	0.688
bremerberta	0.66	0.673
beidazihui	0.8	0.8
bantubvd	0.68	0.618
brichallchapacuran	0.56	0.507
bodtkhobwa	0.78	0.767


Surprise dataset Results for 0.10 data Condition
Mockingbird
Our Model
BLEU Scores
Wangbai
Luangthongkumkaren
Kesslersignificance
Hillburmish
deepadungpalaung
bremerberta
beidazihui
bantubvd
brichallchapacuran
bodtkhobwa
Language sets