

Subject **Burushaski Language Project**
From <chris@christianbentz.de>
To <sadaf.munshi@unt.edu>
Date 2019-04-11 11:42



-
- myh_con_1 (3 KB)
-

Dear Prof. Munshi,

my name is Chris Bentz, I'm a PostDoc at the University of Zurich involved in a project to build a text corpus for 100 typologically diverse languages in order to also give under-resourced languages a place in corpus linguistics (<https://www.spur.uzh.ch/en/departments/research/textgroup/MorphDiv.html>). In our search we came across your Burushaski Language Corpus, which is of course a fantastic resource! We would be interested to transform the pdf files into a plain text format with standardized annotation to make automatic processing possible, see an example file of stories in another language (Makah) attached.

We also have some funding and expertise available to help with transcribing hand-written field notes, so this could be mutually profitable. Our long-term aim is to make the 100 language corpus available under a CC by 4.0 (non-commercial) license to the public.

Let me know if you are interested to collaborate,
Thanks for your time,
Chris Bentz