

# LINGUISTIK



## Zusammenstellung von Korpora

Bisher wurden Ressourcen in den Sprachen Kamassisch, Selkupisch, Dolganisch und Ewenkisch zusammengestellt. Diese beinhalten unter anderem Sprachaufnahmen und Feldforschungsnotizen. Zudem wurde eine umfangreiche Forschungsbibliographie erstellt.

## Annotation und Glossierung

Die Korpusdaten werden glossiert und weitreichend annotiert. Zu allen Korpora wird hierzu eine ausführliche Dokumentation erstellt.

## Grammatische Analysen

Sie bilden die Grundlage für grammatische Analysen und weiterführende interdisziplinäre Forschungen.

## Grammatiken, Korpora und Sprachtechnologie für

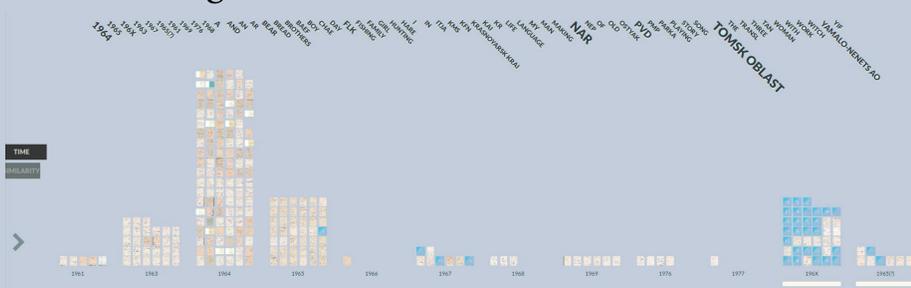


indigene nordeurasische Sprachen  
(Indigenous Northern Eurasian Languages)

## DIGITALE

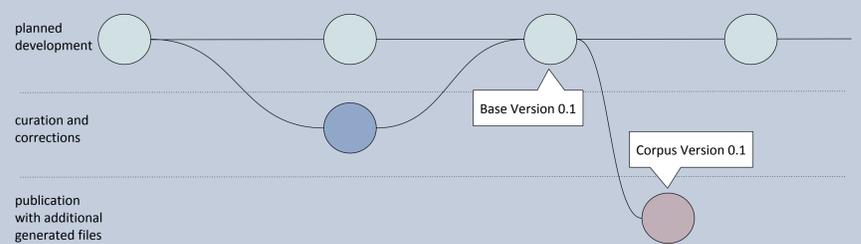
### Forschungsdatenmanagement

Zum kollaborativen Arbeiten wurden Versionierungs- und Kuratierungsroutinen implementiert, die sowohl eine stetige Arbeit an den Daten, als auch eine stetige automatische Überprüfung und Überarbeitung der Daten ermöglichen.



### Open Access Publikation

Bisher wurden Kamas Corpus 1.0, Selkup Corpus 1.0 und Dolgan Corpus 1.0 mit umfangreichen Annotationen und Metadaten veröffentlicht. Diese Daten sind FAIR, also auffindbar, zugänglich, interoperabel und wiederverwendbar.



## Visualisierungen und Analyse

Als weitere digitale Zugangsformen, die zusätzliche Analysen ermöglichen, werden unterschiedliche Visualisierungen möglichst automatisiert erstellt und die hierzu entwickelte Software ebenfalls zur Verfügung gestellt.

