KorporaundKorpustechnologieimProgrammbereich„MündlicheKorpora“amIDS

ImProgrammbereich„MündlicheKorpora“derAbteilungPragmatikdesInstitutsfürDeutscheSpracheMannheim(IDS)werdenKorporagesprochenerSpracheerstelltundarchiviert(FOLKundAGD)sowieWerkzeugezurErstellungundBearbeitungvonKorporaentwickelt(FOLKERundOrthoNormal).DerPublikationvonKorporaundderRecherchedarindientdieDatenbankfürGesprochenesDeutsch(DGD2).FürdieDokumentationvon„BestPractice“-VerfahrenzurKorpuserstellungundfüreinHandbuchzurArbeitmitmündlichenSprachkorporastehtdasGesprächsanalytischeInformationssystem(GAIS)zurVerfügung.

ArchivfürGesprochenesDeutsch(AGD)

DasArchivfürGesprochenesDeutsch(AGD)istdiezentraleSammelstellefürKorporadesgesprochenenDeutsch.DieKorporawerdenimAGDaufbereitetundfürForschungundLehrebereitgestellt.ImAGDwerdenTonaufnahmendigitalisiertundmaskiert,DokumentationsdatennacheinemeinheitlichenSchemaerfasstundTranskripteinnachhaltignutzbareFormatekonvertiert.AusdemGesamtbestandvonüber40Korporasindderzeit17Bestandskorpora(mitca.9000dokumentiertenEreignissen)fürdieDGD2fertiggestellt.Aktuellwerdenim AGDz.B.Korporazu„Jugendsprache“und„DeutschinAustralien“bearbeitet.UnterderAdresse agd@ids-mannheim.debeantwortetdasAGDServiceanfragenzuseinenBeständenundzurKorpustechnologie.

Web: http://agd.ids-mannheim.de

Kontakt: agd@ids-mannheim.de

Forschungs-undLehrkorpusGesprochenesDeutsch(FOLK)

MitdemForschungs-undLehrkorpusGesprochenesDeutsch(FOLK)bautdasIDSeinkontinuierlichwachsendesKorpusauf,umGesprächsdatenausunterschiedlichenBereichendesgesellschaftlichenLebens(Arbeit,Freizeit,Bildung,Medien)imdeutschenSprachraumviaInternetzugänglichzumachen.DieDatenakquiseumfassteigeneAufnahmen,eineÜbernahmeausanderenIDS-Projekten(z.B.„MapTask“-AufnahmenundInterviewsaus„Deutschheute“),DatenspendenausexternenProjekten(u.a.von„GesprocheneWissenschaftssprachekontrastiv“(GeWiss)und„SprachvariationinNorddeutschland“(SiN)),sowieRundfunk-Mitschnitte.BeiderDatenaufbereitungwerdenAufnahmenundTranskriptemaskiert,dieMetadatennachDGD-Schemaeingegeben,TranskriptionennachcGAT-KonventionenfürdasMinimaltranskriptmitdemEditorFOLKERerstelltundieTranskriptemitOrthoNormalorthographischnormalisiertsowiemitdemTreeTaggerlemmatisiert.99Ereignisse(ca.70Aufnahmestunden)sindfürdieDGD2aufbereitetworden;weitere110Aufnahmestundenliegenvorundwerdenbearbeitet.

Web: http://agd.ids-mannheim.de/folk.shtml

FOLKERundOrthoNormal

FOLKER,AkronymfürFOLK-EditoR,istalsTranskriptionseditoroptimiertfürdieArbeitinFOLK,mitGATundmit FOLK-Daten.FOLKERbasiertaufEXMARaLDAundisthiermitsowiemitPraat,ELANetc.interoperabel.FOLKER stehtzumkostenlosenDownloadbereit;seitMärz2009gibtesüber2500Registrierungen.FOLKERwirdinder Version1.2eineSyntaxkontrollefürGAT-Minimal-undBasistranskripteanbieten;dafürwirdderzeitschonein Previewangeboten.OrthoNormalisteinToolzurorthographischenNormalisierunginFOLK,allgemeiner:zum manuellenAnnotierenaufTokenebene.

Web: http://agd.ids-mannheim.de/folker.shtml

DatenbankfürGesprochenesDeutsch(DGD2)

DieDatenbankfürGesprochenesDeutsch(DGD2)präsentierteineAuswahlausdenAGD-Korporaundden gesamtenFOLK-Bestand,ermöglichteinBrowsinginMetadaten,Aufnahmen,Transkripten,Zusatzmaterial sowiemehrereRecherchemöglichkeiten:eineVolltextrechercheinMetadatenundTranskriptensowieeine tokenbasierteRechercheinTranskripten,jeweilsmitRückgriffaufTranskripteundAufnahmen.DieDGDbietet zudemeinenDownloadausgewählterkompletterDatensätze.

Web: http://dgd.ids-mannheim.de