



HAL
open science

Parallèles et dépendances : des corpus inédits pour explorer les dialectes alsaciens

Delphine Bernhard, Nathanaël Beiner, Barbara Hoff, Carole Werner

► **To cite this version:**

Delphine Bernhard, Nathanaël Beiner, Barbara Hoff, Carole Werner. Parallèles et dépendances : des corpus inédits pour explorer les dialectes alsaciens. Humanités numériques en pédagogie et recherche 3, Université de Strasbourg, Feb 2026, Strasbourg, France. 2026, <10.34847/NKL.7DEFAN31>. <hal-05540385>

HAL Id: hal-05540385

<https://hal.science/hal-05540385v1>

Submitted on 6 Mar 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-SA 4.0 - Attribution - Non-commercial use - ShareAlike - International License

Parallèles et dépendances : des corpus inédits pour explorer les dialectes alsaciens

Delphine Bernhard, Nathanaël Beiner, Barbara Hoff, Carole Werner

Corpus parallèle alsacien – français | corse | occitan | poitevin-saintongeais

- Textes alignés entre toutes ou une partie des langues :
 - ~100 000 mots alsaciens :
 - 45% alsacien langue originale, 55% traductions du français
 - Domaines : littéraire, juridique, religieux, journalistique
- Consultation en ligne sur le site ParCoLab et corpus de pièces de théâtre téléchargeable sur Nakala



Corpus théâtral parallèle pour l'alsacien

français

Alphonse Daudet: Contes du lundi (extraits) — (DIVITAL) ©

Un moment l'idée me vint de manquer la classe et de prendre ma course à travers champs. Le **[temps]** était si chaud, si clair ! On entendait les merles siffler à la lisière du bois, et dans le pré Rippert, derrière la scierie, les Prussiens qui faisaient l'exercice.

Gustave Stoskopf: Mossieu le Maire — 1922 ©

Toujours, toujours! Mais, depuis quelque **[temps]**, elle vieillit beaucoup.... Ainsi, par exemple, — soit dit entre nous, elle a oublié de faire, pour la fête, des pâtés au jus et des boulettes de moëlle, de plus, elle n'a fait cuire qu'une livre de saucisse par personne.

alsacien

Alphonse Daudet: Maandigschichtle — 2023 (DIVITAL) ©

Kürz isch mr ingfalle, ich kénnt d'Schuel schwanze ùn d'ürich de Raps gehn. S **Watter** isch so wàrm gewann, so hall! Mr het d'Àmsle àm Wàldrànd heere pfiife ùn uff im Rippert sinere Màtt h'inter de Sajeréi d'Preisse heere iewe.

Gustave Stoskopf: D'r Herr Maire — 1898 ©

Allewyl, allewyl! — Nur in d'r letschte **Zytt** isch se n'ne bissel alt worre.... Hytt hett se zuem Exempel — unter ùns g'sait — vergasse, uff de Maßti Suppepastetle ze mache un Marikknöpfle, un het numme àn Pfund Brotwurscht iwer gemacht uff de Mann.

poitevin-saintongeais

Alphonse Daudet: Contes du lundi (extraits) — 2024 (DIVITAL) ©

In moument l'idàie me vénghit de mancàe la classe é de prenre ma corse a chanp travérs. Le **tenp** étét si chàud, si clér ! I entendiun lés marles subllàe a l'eùràie dau boes, pi den le prai Ripért, dare la sejherie, lés Prusséns çhi fasiant l'émenajhe.

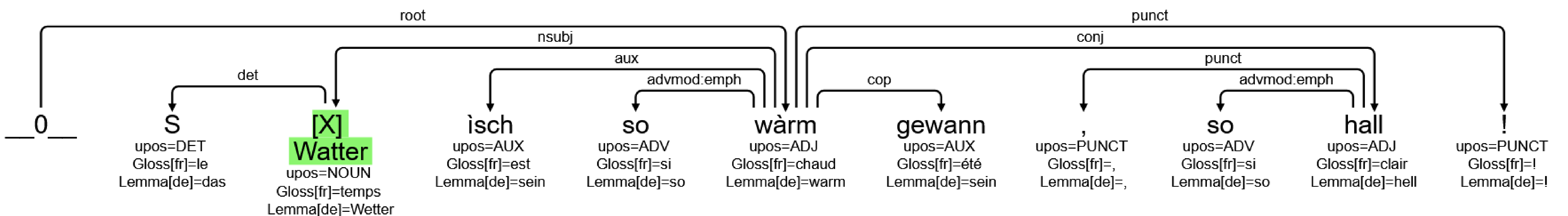


Corpus annoté syntaxiquement

- Premier corpus annoté en relations de dépendances syntaxiques pour les dialectes alsaciens (alémanique) et guide d'annotation
- Correction manuelle après pré-annotation automatique
- Taille : 977 phrases, 19 743 mots



Corpus UD_Alemannic-DIVITAL



Utilisations

- Recherche en linguistique (analyses contrastives, syntaxe)
- Apprentissage et enseignement des dialectes alsaciens
- Développement d'outils de traitement automatique des langues