

tei-corpo

Conversion tool from Elan, Clan, Transcriber and Praat files to TEI files and back

Java library and Swing user interface

Les conversions peuvent être faites en ligne à cette adresse : <http://ct3.ortolang.fr/teiconvert/>

L'outil Java avec interface utilisateur de conversion de formats (TEI_CORPO, CLAN, ELAN, Transcriber, Praat) peut être téléchargé ici : [conversions.jar](#)

Attention : il faut avoir installé Java sur son ordinateur : [Télécharger Java](#)

Utilisation de l'outil de conversion

L'outil est à utiliser directement depuis le Finder ou Bureau (faire double clic sur le nom de fichier).

L'outil est utilisable en ligne de commande. Les commandes à exécuter sont :

- Chat -> TEI_CORPO :
 - `java -cp conversions.jar fr.ortolang.teicorpo.ClanToTei [paramètres]`
- Transcriber -> TEI_CORPO :
 - `java -cp conversions.jar fr.ortolang.teicorpo.TranscriberToTei [paramètres]`
- Praat -> TEI_CORPO :
 - `java -cp conversions.jar fr.ortolang.teicorpo.PraatToTei [paramètres]`
- Elan -> TEI_CORPO :
 - `java -cp conversions.jar fr.ortolang.teicorpo.ElanToTei [paramètres]`

Toutes les commandes utilisent les mêmes paramètres d'entrée sortie:

- `-i` nom du fichier ou répertoire où se trouvent les fichiers à convertir
- `-o` nom du fichier de sortie des fichiers ou répertoire des fichiers résultats

Si l'option `-o` n'est pas spécifié, le fichier de sortie aura le même nom que le fichier d'entrée, avec une autre extension, et sera stocké au même endroit

La conversion depuis Praat dispose de paramètres supplémentaires

- `-p` fichier_de_paramètres: contient les paramètres sous leur format ci-dessous, un jeu de paramètre par ligne.
- `-m` nom/adresse du fichier média
- `-e` encoding (par défaut detect encoding)
- `-d` default UTF8 encoding
- `-t` tiername type parent (descriptions des relations entre tiers)
 - types autorisés: `- assoc incl symbdiv timediv`

- TEI_CORPO -> Clan :
 - java -cp conversions.jar fr.ortolang.teicorpo.TeiToClan [paramètres]
- TEI_CORPO -> Transcriber :
 - java -cp conversions.jar fr.ortolang.teicorpo.TeiToTranscriber [paramètres]
- TEI_CORPO -> Elan :
 - java -cp conversions.jar fr.ortolang.teicorpo.TeiToElan [paramètres]
- TEI_CORPO -> Praat :
 - java -cp conversions.jar fr.ortolang.teicorpo.TeiToPraat [paramètres]
- TEI_CORPO -> Txm :
 - java -cp conversions.jar fr.ortolang.teicorpo.TeiToTxm [paramètres]
- TEI_CORPO -> Lexico :
 - java -cp conversions.jar fr.ortolang.teicorpo.TeiToLexico [paramètres]

Toutes les commandes de conversion inverse utilisent les mêmes paramètres, en plus des paramètres -i et -o

- -p fichier_de_paramètres: contient les paramètres sous leur format ci-dessous, un jeu de paramètre par ligne.
- -n niveau: niveau d'imbrication (1 pour lignes principales)
- -a name : le locuteur/champ name est produit en sortie (caractères génériques acceptés)
- -s name : le locuteur/champ name est supprimé de la sortie (caractères génériques acceptés)

Paramètres supplémentaires pour les exports vers Txm et vers Lexico

- -tv "type:valeur" : un champ type:valeur est ajouté dans les <w> de txm ou lexico ou le trameur
- -cleanline : produit des fichiers sans codes spécifiques de transcription orale
- -section : ajoute un indicateur de section en fin de chaque énoncé (pour lexico/le trameur)

Commandes supplémentaires pour éditer automatiquement les fichiers TEI

- java -cp conversions.jar fr.ortolang.teicorpo.TeiEdit [paramètres] Cette commande permet de modifier les valeurs des champs media, mediamime, docname et les valeurs temporelles dans la timeline Pour cela utiliser l'option -c commande=valeur
 - -c media=nom_de_fichier
 - -c mediamime=valeur
 - -c docname=nom-de_fichier (nom interne du document utilisé pour l'interrogation en xml)
 - -c chgtime=valeur (décale tous les repères temporels de 'valeur')
 - -c replace (ne crée pas un nouveau fichier mais remplace l'ancien)

Il est possible de télécharger la DTD: [DTD TEI Oral](#)