

Le corpus

Notre corpus se compose de 35 dictées issues d'un concours d'orthographe populaire en Belgique francophone, *La dictée du Balfroid*, créé en 1988. Ce concours est accessible à tous les enfants en fin de scolarité primaire. Les inscriptions se font soit via l'école, soit de manière individuelle, avec un maximum de 4 élèves par classe. Chaque année, six demi-finales permettent de sélectionner les élèves qui peuvent accéder à la finale. Ceux-ci ne peuvent avoir fait plus de 5 fautes lors de l'épreuve éliminatoire pour laquelle aucune sélection n'est effectuée.

Nous avons constitué une base de données, construite selon les procédures décrites dans Dister et Moreau (2017) ([***lien](#)).

Le corpus se compose des dictées des années 2012 à 2016. Nous avons attribué un code à chaque dictée. Celui-ci reprend l'année de la dictée, une lettre pour indiquer s'il s'agit d'une demi-finale (D) ou d'une finale (F), ainsi qu'un substantif qui figurait dans le titre originel. Ces 35 dictées comptabilisent 11 298 copies, avec une moyenne de 132 mots par dictée.

Le tableau ci-dessous synthétise les informations pour chaque dictée.