A Diachronic Digital Edition of the Petit Larousse illustré

Herve Bohbot, Alexandre Fauchere, Francesca Frontini, Agata Jackiewicz, Giancarlo Luxardo, Agnès Steuckardt, Mohamed Khemakhem, Laurent Romary

To cite this version:

HAL Id: hal-01873805
https://hal.archives-ouvertes.fr/hal-01873805
Submitted on 13 Sep 2018

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.
The Petit Larousse illustré is the only French dictionary to be updated every year, since it was first published in 1905 (coined 1906).

Handy and affordable, with each edition selling several hundreds of thousands copies (Pruvost, 2002), one Petit Larousse may virtually be found in every French household.

Editions published prior to 1948 are public domain according to French laws, however, no open full text resource is currently available.

WEB site

Targeting both scientific (linguists, historians, lexicographers, social scientists) and general public. Simple and advanced research interface, responsive design.

Public release in July 2018 with 1906 to 1924 editions, then new data will be regularly added.

TEI XML

Reference digital edition format. For long-term preservation and expert querying. Users: digital humanists, computational linguists. The use of TEI-LEX0 format (Banski et al., 2017) will allow for a predictable structure and for the exploitation of ad hoc digitisation technologies such as GROBID-Dictionaries (Khemakhem et al., 2017)

Ontolex-Lemon RDF

Natural language processing applications and linking to other resources.

Users: language technologists, computer scientists.

A new resource for studying recent history of language, culture and techniques

An example with the entry AVIATION between 1906 and 1912 editions

Some word datations:

1906: tract
1910: gribiste (striker)
1911: antimilitarisme
1912: puériculture, technicien
1914: tract
1915: copyright
1916: scientisme, syndicalisme
1918: boche (pej. for German)
1920: alerter
1921: bolchévisme
1922: mazout, pogrom, soviet
1923: bolschivy
1924: bolchévik
1925: bolchévisme
1929: bolchevique
1930: Bolchevisme
1934: bolchevisme
1939: bolchevik
1942: bolchevisme
1944: bolchevisme
1946: bolchevisme
1947: bolchevisme
1948: bolchevisme
1950: bolchevisme
1955: bolchevisme
1960: bolschivy
1964: bolschivy
1966: bolschivy
1970: bolschivy
1975: bolschivy
1980: bolschivy
1985: bolschivy
1990: bolschivy
1995: bolschivy
2000: bolschivy
2005: bolschivy
2010: bolschivy
2015: bolschivy
2020: bolschivy
2025: bolschivy

References:


Nénufar project is headed by laboratoire Praxilting at Paul-Valéry University of Montpellier in collaboration with Inria. It is supported by the Délégation Générale à la Langue Française et aux Langues de France (an agency of French Culture and Communication Ministry) and the Huma-Num consortia CORLI and CAHIER.