

## French Heterophonic Homographs

The descriptions in this document are extracted from [Hajj et al., 2022].

**Definition** An heterophonic homograph is “one of two or more words spelled alike but different in meaning or pronunciation”<sup>1</sup> (such as ‘mon fils’ (my son) vs. ‘des fils’ which is the plural of ‘fil’, a thread or wire). The correct classification of a homograph has always been an issue for natural language processing when it comes to text-to-speech systems and the analysis of texts. And so, using a word that has homographs, can change the meaning and context of a text if read or analysed by a machine. There are 803 heterophonic homograph in French (source and list: [https://fr.wiktionary.org/wiki/Catégorie:Homographes\\_non\\_homophones\\_en\\_français](https://fr.wiktionary.org/wiki/Catégorie:Homographes_non_homophones_en_français)).

**Examples** The following examples are given with the phonetic alignment format introduced in the Blizzard 2023 training dataset [https://zenodo.org/record/7293056#.Y7wkRC\\_pMzw](https://zenodo.org/record/7293056#.Y7wkRC_pMzw)

- Adoptions [adɔpsjɔ̃] / [adɔptjɔ̃]  
 .Pourquoi les **adoptions** internationales s’effondrent? [adɔpsjɔ̃]  
 \_\_p\_u\_r\_k\_w\_a\_l\_e\_z\_a\_do^psj\_o~\_z\_e~\_te^r\_n\_a\_s\_j\_o^na\_l\_---s\_e\_fo~\_drq\_---  
 .Je voulais que nous **adoptions** la même approche. [adɔptjɔ̃]  
 \_\_z^x^\_v\_u\_l\_e^---k\_x^\_n\_u\_z\_a\_do^ptj\_o~---l\_a\_m\_e^m\_\_a\_p\_r\_o^s^---
- Excellent [ɛksɛlã] / [ɛksɛl]  
 .Il parle un **excellent** français. [ɛksɛlã]  
 \_\_i\_l\_p\_a\_r\_l\_---x~\_n\_e^k\_s\_e\_l\_a~---f\_r\_a~\_s\_e^---  
 .Il faut travailler avec des gens qui **excellent**. [ɛksɛl]  
 \_\_i\_l\_fo---trava\_\_j\_e\_r\_a\_v\_e^k\_d\_e---z^a~---k\_i\_e^k\_s\_e^\_l\_---
- Bus [bys] / [by]  
 .Elle court pour ne pas rater son **bus**. [bys]  
 \_\_e^\_l\_---k\_u\_r\_\_p\_u\_r\_n\_x^\_p\_a\_\_r\_a\_t\_e\_\_s\_o~\_\_b\_y\_s\_\_  
 .Les jus de fruits doivent être **bus** avec modération. [by]  
 \_\_l\_e\_\_z^y\_\_d\_x^\_f\_r\_h\_i\_---d\_w\_a\_v\_\_t\_e^t\_r\_q\_b\_y\_z\_a\_v\_e^k\_m\_o^d\_e\_r\_a\_s\_j\_o~---

### Reference

[Hajj et al., 2022] Hajj, M.-L., Lenglet, M., Perrotin, O., and Bailly, G. (2022). Comparing NLP Solutions for the Disambiguation of French Heterophonic Homographs for End-to-End TTS Systems. In Prasanna, S. R. M., Karpov, A., Samudravijaya, K., and Agrawal, S. S., editors, *Speech and Computer*, pages 265–278, Cham. Springer International Publishing.

<sup>1</sup><https://www.merriam-webster.com/dictionary/homograph>