

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”¹ (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).

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

- 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.

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