

Transcriptions MPF

version finale janvier 2019

Nombre de tires	Autant que de locuteurs + 2 (1 tire silence et 1 tire commentaire)
Orthographe	Globalement, orthographe standard - Pas de trucage de type <i>j'sais, chais, ya, t'as</i> (faire spécialement attention à <i>tu as, tu es, qui a, il y a</i>), ... - Pas d'API - Les mots sans orthographe standardisée doivent suivre au maximum les contraintes de l'orthographe standard. Voir le comité lexique, qui a fixé plusieurs mots (arabe, verlan, argot).
Amorces	Un tiret. On ne le met que si le mot est clairement incomplet : <i>ta-tab- table</i> Ne pas mettre de tiret quand un mot est prononcé sans être terminé, mais pas prononcé complètement ensuite.
Silences, tire silence et pauses	Une tire « Silences » spécialement dédiée règle le problème de savoir à qui attribuer les silences, du fait qu'il est indécidable si c'est le locuteur en place qui se tait, ou un autre qui ne prend pas la parole. Une pause brève est notée (.), une longue (..). De préférence dans la tire Silences, mais peut être interne à un tour de parole s'il est clair que le locuteur ne s'interrompt pas.
Début et fins d'énoncé	Les énoncés commencent tous par une majuscule et se terminent d'une des façons suivantes : - Un point - Un point d'interrogation pour une question - Un point d'interrogation entre parenthèses pour une question supposée La ponctuation finale est toujours à la toute fin d'un énoncé (donc, après l'éventuel chevron fermant).
Segments inaudibles et indécision	Marqués par autant de x que de syllabes perçues/perceptibles. Il vaut mieux mettre (xxx) que forcer l'interprétation quand on n'est pas sûr d'entendre correctement.
Chevauchements	Les segments superposés sont placés entre chevrons (ils vont toujours par 4, deux ouvrants et deux fermants) L1 : <i>yyyyuiio<xx>oioioi</i> L2 : <i><tyty></i> Les chevrons doivent entourer des syllabes entières, pas des bouts
Bruits de bouche	Indiqués entre parenthèses dans le texte lorsqu'il y a une rupture sensible dans le discours Ex. : (baillement), (clic), (inspiration)
Commentaires et perturbations occasionnels	Figurent dans la tire « Commentaire », entre parenthèses (<i>L répond au téléphone</i>), (<i>Bruit de voiture</i>), (<i>Sonnerie de téléphone</i>), (<i>Train qui passe</i>)...

Transcriptions MPF

version finale janvier 2019

Mots étrangers et mots non standard	En texte, le mot est suivi de @s et commenté/traduit dans la tire commentaire En texte : <i>shit@s, meuf@s</i> portée : lorsque le @s concerne plusieurs mots, sa portée est indiquée avec des [et des]. Par exemple <i>[naal sheytan]@s, [mal au trou]@s, [salam alikoum]@s, [frères mus]@s, [trou du cul]@s</i>. Dans la tire « Commentaire » : &=ver femme (schéma strict) la catégorie est donnée par les 3 premières lettres de la langue ou du procédé. Langues : &=ang (anglais), &=ara (arabe), &=rom (romani), &=nou (nouchi) Procédés : &=ver (verlan), &=tro (troncation), &=sig (sigle), &=apo (apocope), &=aph (aphérèse) Catégories : &=arg (argot), &=enf (langage enfantin), &=fns (français non standard). Cette dernière catégorie a été ajoutée pour ce qui n'est pas de l'argot héréditaire
Anonymisation	Seuls les pseudos apparaissent dans la transcription (noms des tires et dans le texte). Les pseudos sont suivis de @a, ex : Jean-Marc (nom de l'informateur) est transcrit <i>Stéphane@a</i> (pseudo@a). Après anonymisation du son, tout @a doit avoir disparu.
Mots épelés, acronymes	Un seul mot quand c'est syllabé, comme <i>GAV</i> pour Garde à Vue, et lettres détachées quand c'est épelé, comme <i>G A V</i> (3 mots). On peut donc en théorie avoir les deux notations (parce que les deux prononciations) dans la même enquête.

Transcription des onomatopées et autres petits mots...

(on ne cherche pas à reproduire la nuance précise. Majuscule seulement si en début de séquence)

ah
beh
beuh
etc
hum hum (acquiescement)
oh
ouf
psst
chut

aïe
ben
eh bien
euh
mh
ouah
ouh
ts
gnagnagna

bah
beurk
mouais
hum
ouais
ok
pff
blablabla
nananananane

Transcriptions MPF

version finale janvier 2019

Conseils aux transpositeurs sur ce à quoi il faut veiller

élaborés à mesure de la transcription du corpus

- *les pièges de transcription les plus fréquents qu'il faut veiller à éviter*

il y a transcrit y a (quand prononcé [ja])

tu as ou tu es transcrit avec une apostrophe quand il y a élision du u de tu

dans les deux cas, ces transcriptions ne sont pas bonnes, puisqu'on a décidé 1) de toujours garder l'intégrité du mot (le mot est présent ou absent, mais il n'est pas réduit) ; 2) de noter, dans les cas indécidable, la version standard

- *pour les frontières*

attention, pour les commentaires, décaler légèrement la frontière gauche vers la droite, de façon à ce que, dans le document word, le commentaire figure après et non avant l'énoncé commenté.

Il y a parfois des frontières qui ne servent à rien, les retirer à la révision.

- *pour les commentaires*

comme on a changé de format en cours de route, il y a des restes des anciennes formules de commentaires. Le format actuel est :

&=xxx équivalent (pas d'espace entre « & », « = » et « xxx »), espace après le « xxx », et rien de plus que un ou plusieurs « équivalents »

xxx représente les trois premières lettres du procédé ou de la langue à laquelle est fait

l'emprunt. Soit la langue à laquelle il y a emprunt (ara, ang, bam, rom – attention, romani et non tzigane...), soit arg ou fns (argot, français non standard), soit le procédé (ver pour verlan, apo pour apocope, aph pour aphérèse...)

- *la tire Silences*

au début on n'en mettait pas, et c'est trop de travail d'en créer une s'il n'y en pas. Donc, s'il n'y en a pas, gérer au mieux ; s'il y en a une, ne pas oublier les parenthèses autour du/des point(s) : (.) ou (..).

