

# قواعد كتابة الملفات الصوتية للهجات العربية

هذا الملف يحتوي على القواعد المتبعة عند كتابة المحتوى المقابل لملفات صوتية (أو ما يسمى بتفريغ ملفات الصوت)، بهدف توحيد المنهج عبر كل المستخدمين أثناء تطبيق ذلك للهجات العربية (المصرية، المغاربية، الخليجية، والشامية). سيكون هناك قواعد خاصة بكل لهجة إذا كان هناك حاجة لذلك.

**أولاً:** يتم تقطيع ملف الصوت بامتداد "wav" عن طريق برنامج Transcriber لمزيد من المعلومات عن البرنامج، يمكن الاطلاع على الورقة البحثية الآتية:

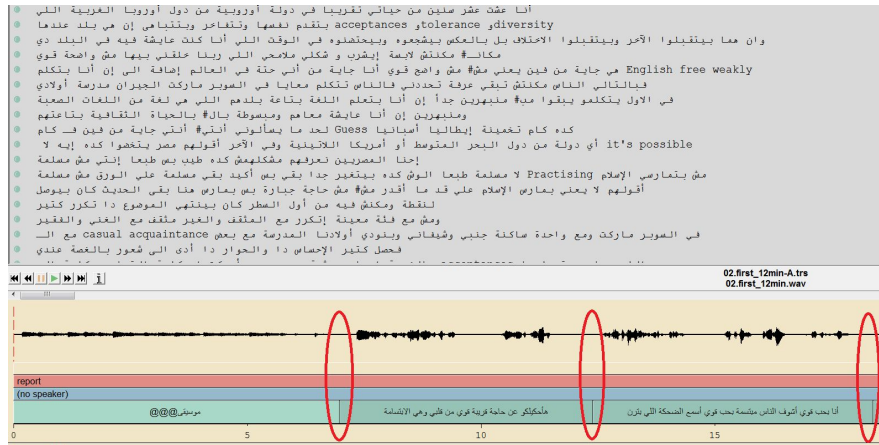
<http://languagelog ldc.upenn.edu/myl/ldc/Transcriber.pdf>

يمكن تحميل البرنامج من الرابط التالي:

<http://trans.sourceforge.net/en/presentation.php>

**ثانياً:** خذ بعض الوقت للتدريب على استخدام البرنامج لتقطيع الصوت، الانتقال بين المقاطع المختلفة، إضافة مقطع جديد، حذف مقطع، كتابة النص المقابل لمقطع ما أو تعديل النص، حذف مقطع وهكذا..  
(فتح ملف الصوت عن طريق القائمة: File/Open audio file)  
ملاحظة: قم بتكبير شاشة البرنامج حتى يظهر أمامك كل شيء!

**ثالثاً:** عند تقطيع ملف الصوت، يجب النظر إلى شكل موجة الصوت ويتم اختيار أماكن التقطيع حتى لا تأتي في منتصف الكلام، كما هو موضح في الصورة التالية، والتي تُظهر أماكن محتملة لبدائيات ونهايات المقاطع الصوتية.



أيضا يجب أن يتراوح زمن كل مقطع في حدود 10 ثواني. لا نلجأ إلى أزمان أكثر من 30 ثانية إلا في حالة الضرورة القصوى. يجب الأخذ في الاعتبار أن تكون نهاية المقطع عند جملة مفيدة، ولا مانع من أن يكون زمن المقطع أقل من ذلك لتحديد نهاية جيدة للمقطع.

**رابعاً:** يجب قبل البدء في كتابة النص أن تختار الترميز UTF-8، عن طريق اختيار الترميز من القوائم الآتية:

Options \ General \ Encoding \ Unicode (UTF-8)

لأنه في حالة عدم اختيار هذا الترميز، سوف يظهر الكلام بطريقة غير مقروءة، وهذا سيتطلب منك إعادة إضافة النص من الجديد ثم اختيار هذا الترميز.

**خامسا:** نظرا لصعوبة تحرير النصوص بالعربية داخل هذا البرنامج، من الأفضل كتابة النص في محرر عربي جيد (مايكروسوفت وورد، أو نوتباد مثلا)، ثم نسخ النص داخل المقطع المطلوب.

**سادسا:** يجب أن تكون الكتابة **مطابقة تماما** لما يتم سماعه (كتابة حرفية)، بمعنى أنه إذا كان هناك تلغثم واضح في الكلام أو تصويب أو تكرار للكلام أو حتى إذا احتوى الصوت على خطأ نحوي مثلا، فيتم كتابة ذلك بكل دقة حتى تُعبر الكتابة تعبيراً حقيقياً عن الصوت. يتم تمييز هذه الكلمات بوضع علامة # مباشرة بعد انتهاء الكلمة.

مثلا: هذا مثال يوضح الكتابة الحرفية المطابقة للصوت المسموع:

**أنا أنا# أريد# أن# نذهب# أريد أن أذهب من ل# المدينة باتجاه الشمال# الجنوب.**

هنا تم تمييز الكلمات المكررة أو التلغثم (أنا، ل)، والكلمات التي حدث عندها تصويب (أريد، أن، نذهب، الشمال) بوضع رمز # بعدها مباشرة.

القاعدة العامة أنه يمكننا بعد حذف هذه الكلمات أن نحصل على سياق مكتمل مترابط: أنا أريد أن أذهب من المدينة باتجاه الجنوب. لحسن الحظ أنه في كثير من الحالات لا يوجد كل هذه الحالات الخاصة! ولكن إذا وجدت فينبغي كتابتها بكل أمانة.

**مثال 1:** الأسلحة بع# أحيانا أو# أو تأمين بعض ال# ال# ال#

**مثال 2:** هذه الحوالة تذهب إلى أشخاص معينين أم إلى عا# عدد من الأهل في# في البلد?... للأهل الوالد والوالدة للأخوة

**سابعاً:** يتم كتابة الأرقام بالحروف العربية وبالطريقة العامية إذا نُطقت بهذه الطريقة.

مثلا: تبلغ النسبة ثلاثة وأربعين وخمستاشر من مية في المية سنويا

**ثامناً:** إذا كانت الكلمة المنطوقة عربية فصيحة فيجب كتابتها بدون أخطاء إملائية (مثلا أخطاء الهمزات، التاء المربوطة، الياء المنقوطة). نلتزم هنا بالياء المنقوطة للترقية بينها وبين الألف المقصورة (مثلا: الهيكل العظمي، الدول العظمي).

**تاسعاً:** في حالات نادرة لا يمكن تمييز الكلام المسموع بدقة - حتى مع تكرار سماع المقطع عدة مرات - من الممكن هنا استخدام كلمة "غير واضح" كبديل لهذه الكلمات.

مثلا: أنا جيمس مايكل من قرية صغيرة اسمها غير واضح غير واضح بجوار مدينة دالاس في تكساس.

**عاشراً:** أحيانا لا تحتوي المقاطع على نص يمكن كتابته (فراغ، أو موسيقى، أو ضوضاء، أو تصفيق مثلا)، في هذه الحالات يتم

كتابة ذلك في بداية المقطع مسبقاً برمز @@@ (مثلا: @@@@ فراغ).

انقر على هذا الرابط للاستماع إلى مقطع تم ترميزه على أنه @@@ موسيقى: [رابط موسيقى](#)

انقر على هذا الرابط للاستماع إلى مقطع تم ترميزه على أنه @@@ فراغ: [رابط فراغ](#)

**أحد عشر:** إذا كان هناك تداخل في أصوات المتحدثين داخل نفس المقطع، بمعنى أنه يصعب تمييز الصوت لأنهم يتحدثون في نفس الوقت، فيتم وضع علامة ##### في بداية المقطع للتنبيه على أنه من المقاطع ذات الصعوبة العالية في التعرف على محتواها. في مثل هذه المقاطع، يتم كتابة ما يتم سماعه قدر الإمكان مع كتابة غير واضح في النقطة التي يتم فيها التداخل.

يمكنك الاستماع للمقاطع التالية التي تحتوي على أصوات متداخلة:

**ملاحظة :** عندما يكون في شخصين يتحدثان (يجب كتابة حديث كل شخص على حدة) وبعد أن ينهي الشخص حديثه يجب إضافة ... (ثلاث نقاط بعد الحديث) ثم تبدأ في كتابة حديث الشخص الثاني وبعدها (يتم إضافة ثلاث نقاط بعد الحديث) وذلك للترقية بين الشخصين

**مثال 1:** #####ضعيف.. أرجو أن تُعيد السؤال.. نحاول أن نُحسنه، نحاول أن نُحسنه، هل هناك تفسير سيد قطبي لظاهرة أن هناك

**مثال 2:** #####ولكن على هذه الدولة أن تمنع الأذى وأن تمنع سياسة إلحاق الضرر.. طيب

**اثنا عشر:** عند الانتهاء من تفرغ النص كاملاً ومراجعته، يتم حفظ الملف عن طريق البرنامج في صيغة trs وهي تحتوي على ما

قمت بكتابته مصحوباً بالزمن المخصص لكل مقطع. بمجرد النقر المزدوج على هذا الملف، سيتم فتح البرنامج مرة أخرى مع

إمكانية الاستماع إلى ملف الصوت وإمكانية تعديل المقاطع إذا أردنا ذلك.

هذا مثال للمنتج النهائي من ملف ال trs وكذلك ملف الصوت: يمكن استخدامها كمرجع.

بعض القواعد الخاصة بكل لهجة على حدة:

● قواعد خاصة للمصري

- عند سماع كلمات أجنبية غير عربية، يتم كتابتها بالحروف العربية إذا كان ذلك مستساغاً، وإلا يتم كتابتها باللغة التي نُطقت بها.
- مثلاً: أطلقت شركة مايكروسوفت برنامج **OneDrive for Business** بالإضافة إلى برنامج إكسيل.
- هاي يا جماعة.. أنا رححت للدكتور وقال لي إن عندك **infection** من فيروس غريب جداً، والموضوع ده طلع **dangerous** قوي.

● قواعد خاصة للخليجي

● قواعد خاصة للمغربي

- في حالات الكلمات الفرنسية التي قد تصل إلى جملة كاملة بالفرنسية، لا بأس بكتابتها بحروف لاتينية.
  - إن كان هناك مزج بين العربية أو الكلمات المعربة، تكتب الجملة بالحروف العربية.
  - المفردات المعربة أي أجنبية لكن وضعت في قالب عربي، تكتب بالحروف العربية:
  - كي نخلصو البروجي هذا نولو نكونفونديو بين العربية و الفغونسي
  - صح نهذرو أفورس دي برا
  - يقولو في الميليو نتاع لي جان، ياكل العسل بودنيه
  - إلا في حالة أن يكون المقطع كاملاً باللغة الأجنبية، فيكتب بالحروف اللاتيني: **Mais jamais j'ai parlé!**
- !de ca

● قواعد خاصة للشامي