英語リスニングスクリプト：テキスト化のための仕様

2015/12/10

# ファイル名

* テキスト化したファイル名は、元のスクリプトPDFの名前と同じにする。名づけ規則は次の通り。
  + {試験名}-{年度（西暦4ケタ）}--{本試験と追試験の区別}-｛科目名｝.txt
  + ハイフンはすべて半角。年度と本試験・追試験別を区切るハイフンは二本（--）とする。

# 文字のテキスト化

* 文字コードはUTF-8で統一する。
* ファイル中の数字、アルファベットはすべて半角とする。丸付き数字（①など）も、半角数字（丸なし）とする。
* 句読点は、日本語部分は「，」（全角カンマ）と「。」を使用する。英語部分は「,」（半角カンマ）と「.」（半角ピリオド）を使用する。
* その他、英語部分の記号はすべて半角とする。
* 英語問題中の英文は、カンマ、ピリオド、コロン、セミコロン、ダブルクオーデーションの後は半角スペース1つ分を空ける。
* 省略を表すアポストロフィの後の半角スペースは不要。
* カンマあるいはピリオドの後にダブルクォーテーションが続く場合は、「"seniors," 」「"yes."」のように、カンマ（あるいはピリオド）の後は半角スペースなし、ダブルクォーテーションの後に半角スペースありとする。
* 丸カッコの全角・半角については以下の通りとする。
  + 文章の中のカッコ書き（番号・記号以外）は、日本語の文章中なら全角、英語の文章中なら半角で統一。

# テキスト化対象外の部分：以下の部分は不要。

* 音声として流れない部分（ファイルの表題、見出し、ページ番号、その他音声として流れない説明部分）

# 音声による確認→追加と書き換え

* 実際の音声を確認し、スクリプトにない音声を追加する。ただし、“Listen again”の後に繰り返して読まれる問題は追加不要。
* スクリプトPDFの英語部分で現れている省略形や数字、記号は、実際に該当する音声を確認して書き換える。

例：

* + 31 → thirty one 50,000 → fifty thousand
  + % → percent
  + $70 → seventy dollars
  + 2:05（時間） → two o five
  + 25th → twenty fifth
  + 1994（年号） → Nineteen Ninety four