lexpro troubleshoot import

Lexique Pro troubleshooting import

I had a mystery this afternoon. A user wanted to import their Toolbox lexicon into Lexique Pro. I went through the import settings dialog, but nothing would import. I verified that the Toolbox file was Unicode, which it was.

Then I thought maybe there was an invalid character in the file somewhere. I opened the file in Word, said it was a UTF-8 text file, then I saved a new copy of it, also as a UTF-8 file. Lexique Pro imported all the content fine from this new file.

I did a binary comparison between the two files, and found: 1) the file I saved from Word had a BOM at the beginning. The absence of the BOM from the first file wasn't the problem. 2) the first file had an invalid character in it. It had a xB4 (Decimal 180, grave accent) in it, which is not a valid Unicode character. Word had removed this character when I saved the second file, so LexiquePro had no problem.

I learned later that there is a check in Toolbox for invalid Unicode characters. It is in the **Checks** menu, **Check Unicode Validity.** But it requires that all language encodings be set to Unicode. If the encoding for English or some other language is not set to Unicode, there could be invalid characters in that encoding, that the check will not find -- but they will still block the import into Lexique Pro.