**Section 1: cat and character encodings**

**Excel:**

1. encoded-excel.xlsx

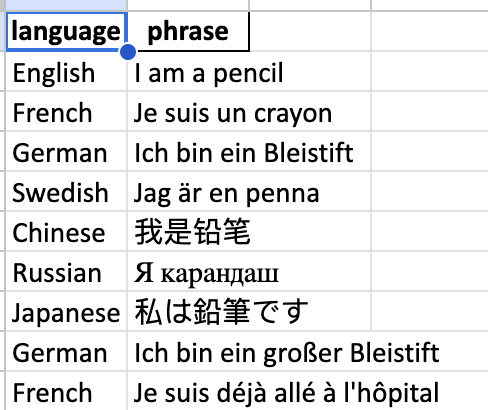


2. encoded-commas.csv

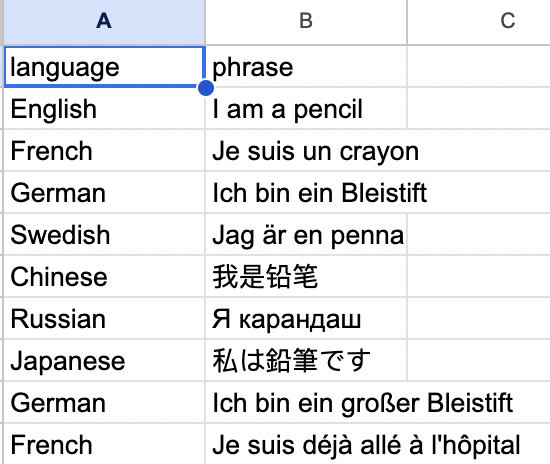


**Google Sheets:**

1. encoded-excel.xlsx



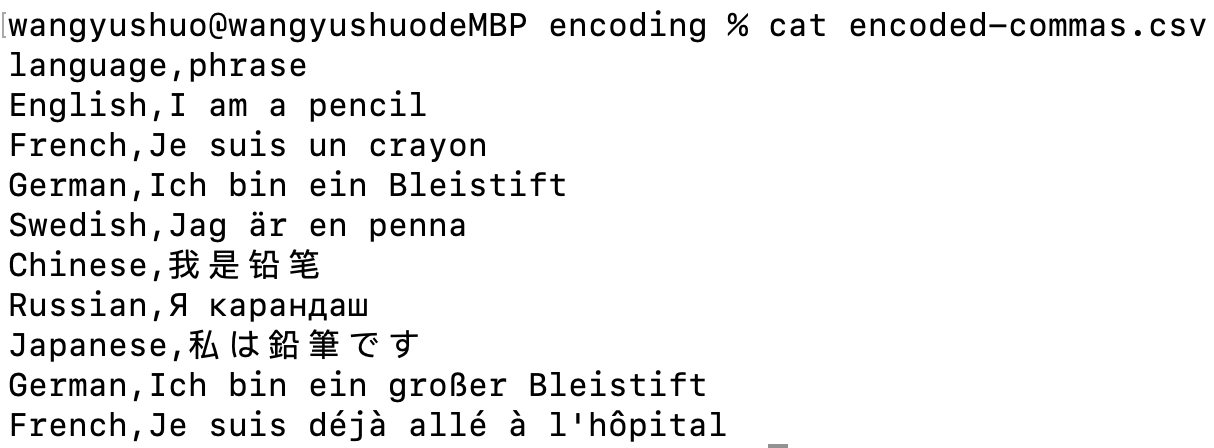
2. encoded-commas.csv



**Cat - Terminal:**

1. encoded-excel.xlsx

2. encoded-commas.csv



**Summary:**

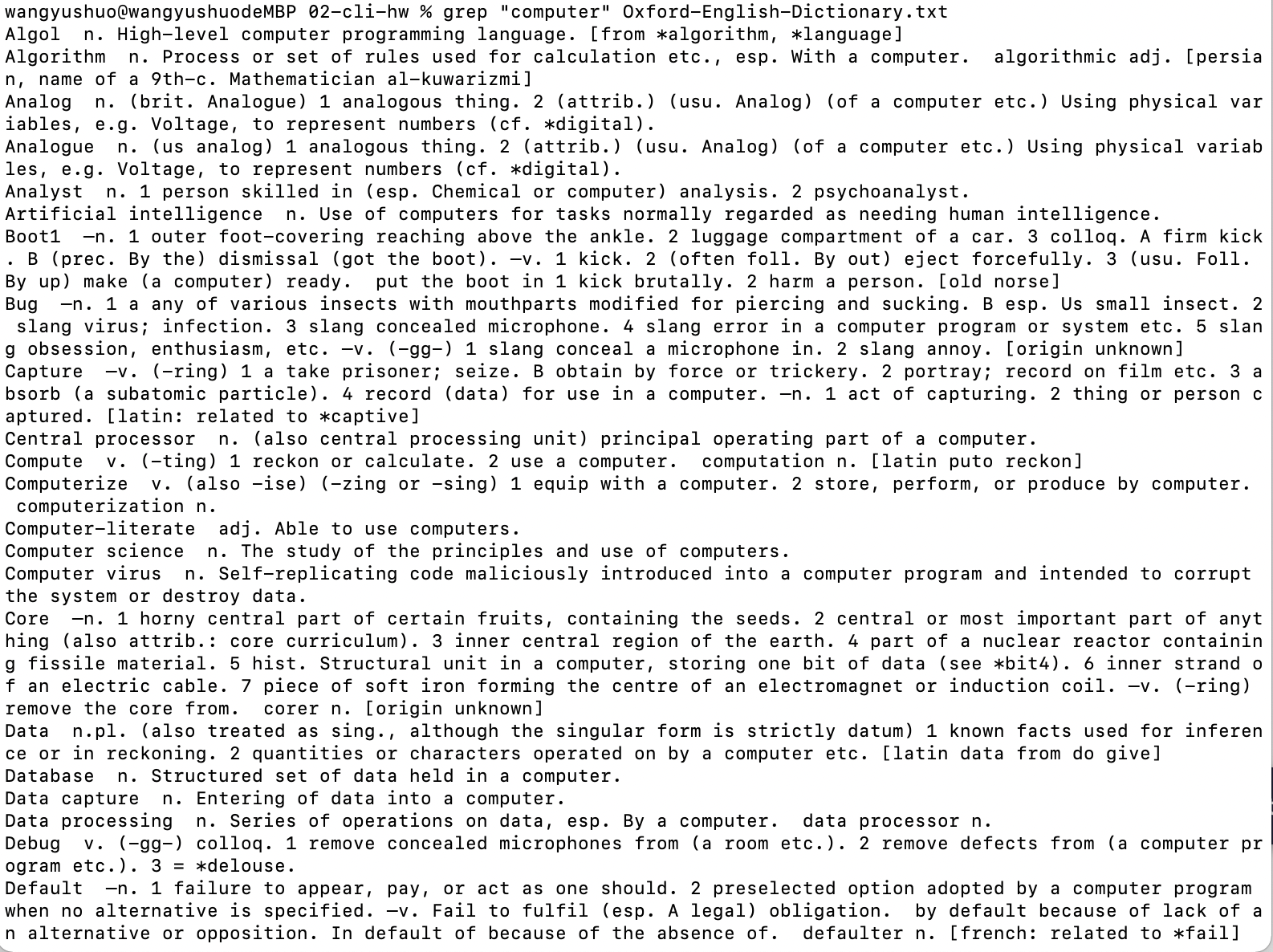
In Excel, encoded-excel.xlsx looks fine for each language, while part of encoded-commas.csv is messed up, especially for languages not using alphabets (Chinese, Russia, Japanese). In Google Sheets, both encoded-excel.xlsx and encoded-commas.csv look fine. In Cat, encoded-excel.xlsx is completely messed up, while encoded-commas.csv, just like the name of “csv” is separated by commas.

**Section 2: grep**

1. Search inside Oxford-English-Dictionary for the word "computer"

Code: grep "computer" Oxford-English-Dictionary.txt

Why: we use grep "literal\_string" filename when we want a particular string in a file

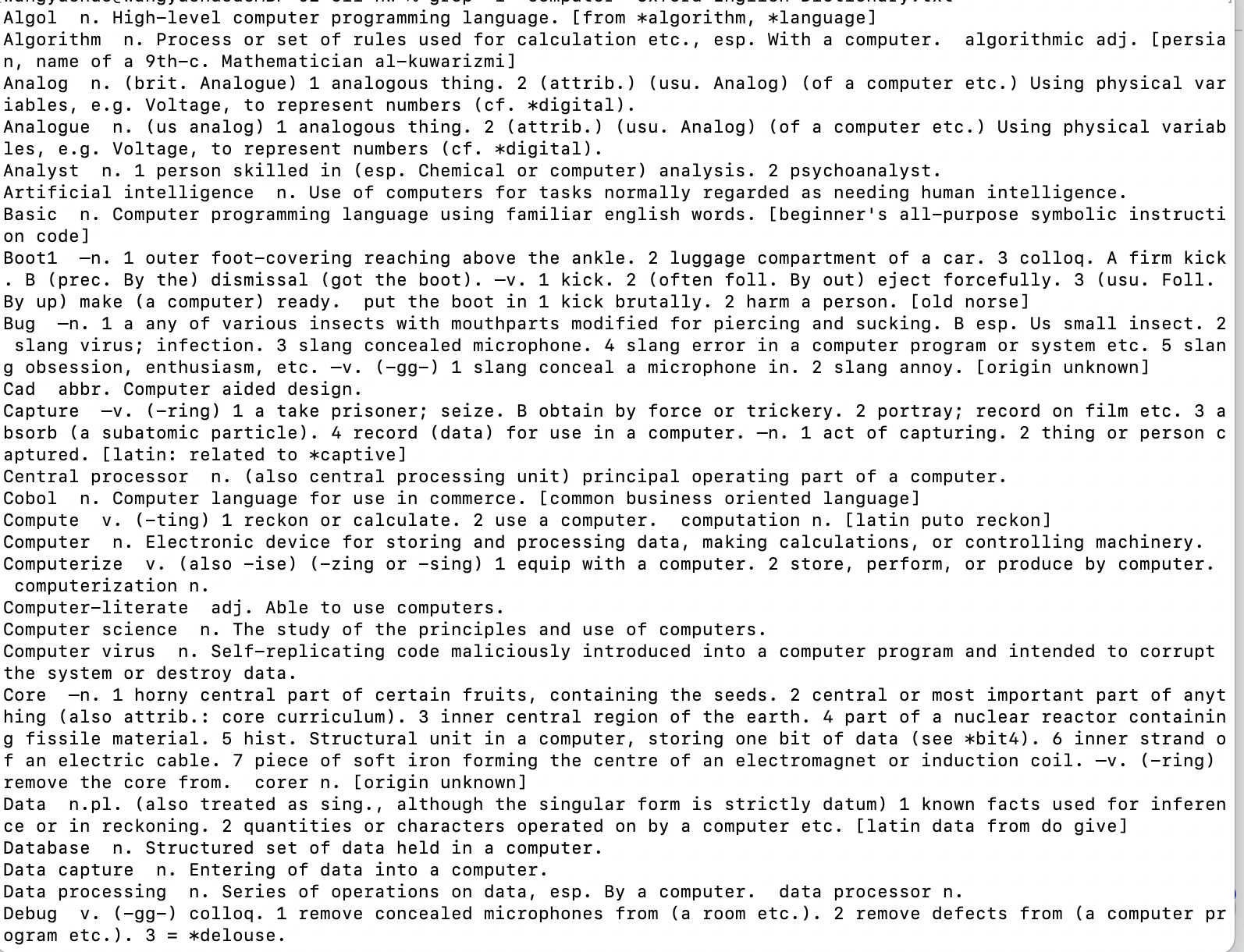


Limitation: this is case-sensitive, so we only get the ones exactly showing lowercase “computer” We miss upper case or other transformations in this way.

2. Figure out what seems wrong with the results. Change it to fix the problem, showing all instances of the word "computer"

Code: grep -i "computer" Oxford-English-Dictionary.txt

Why: we use grep -i "string" FILE that allows us to have all variations of “computer”.

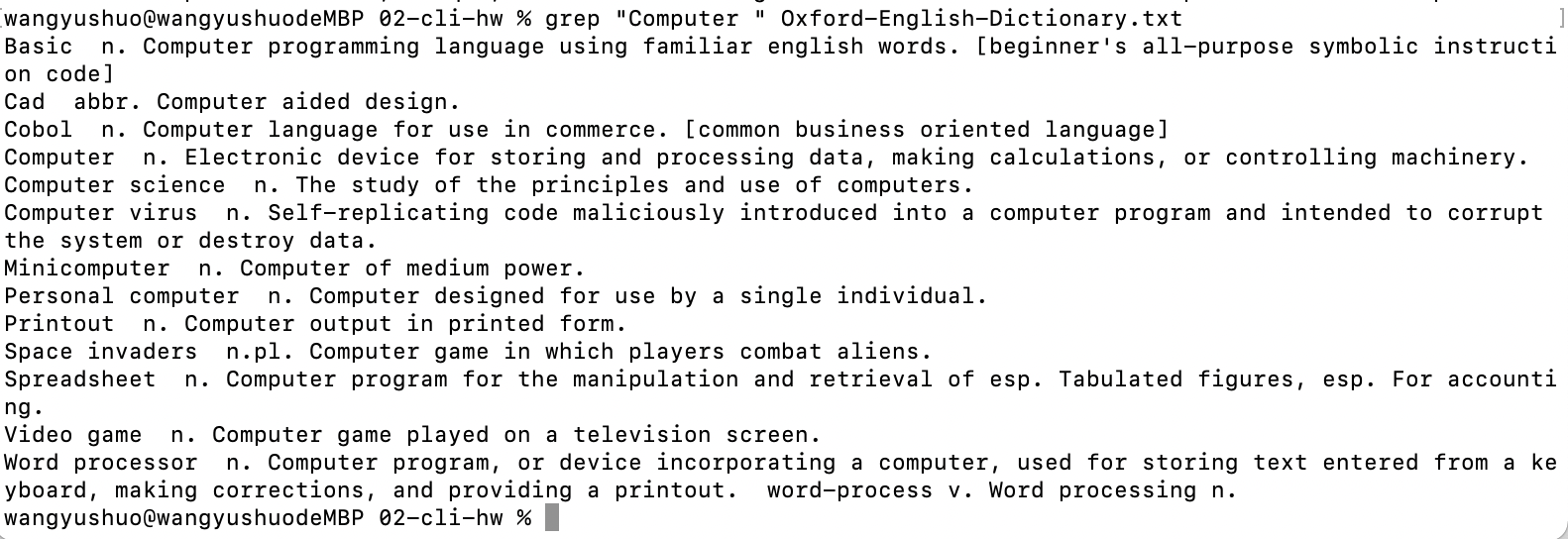


Limitation: it may include too many variations than needed.

3. Let's narrow it down a little: write a grep command to show me Computer but also related phrases like Computer virus and Computer science.

Code: grep "Computer " Oxford-English-Dictionary.txt

Why: the space after Computer will list phrases starting with “Computer” and followed by other words.

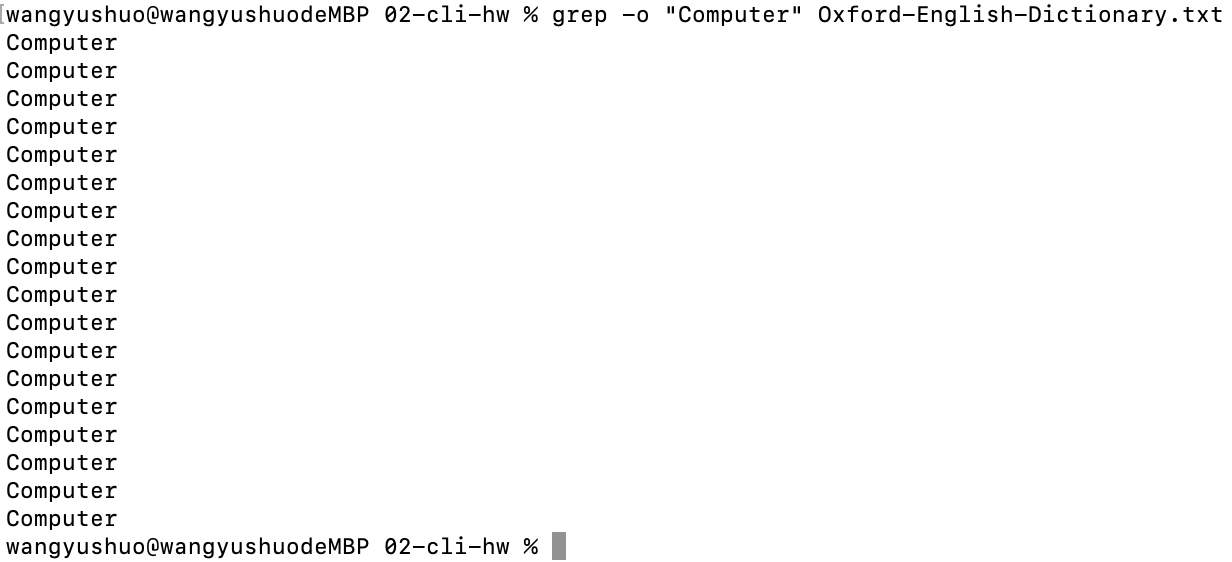


Limitation: we may want to include phrases with lowercase “computer” as well.

4. Finally, we're now only interested in the definition of the word Computer. Write a grep command to show me only that line (not "Computer virus" or "Computer science" etc).

Code: grep -o "Computer" Oxford-English-Dictionary.txt

Why: the -o option helps show out only the matched string of a particular pattern.



Limitation: this lacks context if are too picky on the pattern.