# Unusual cases solved:

## Non-Latin characters:

Texts: 'Grilréishaw, B. E.'; 'Hu§igson. W: H:'; 'Neérﬂnlgrg. A.'; 'Miﬁter, E. M.'

Solution: Update to Python 3.7, and manually add to regular expressions (still need to include more texts manually)

pattern1 = re.compile(**"[A-Z]['‘]\*[a-zé§¢ﬂﬁ]+['‘]\*[a-zé§¢]\*[,.:]\*\s[A-Z]\*\s\*[A-Za-z01][,.:]+\s"**)

## Author family name with a "-"

Texts: "Kaye-Smith, S."

Solution: Include a new case of regular expressions

pattern1\_1= re.compile(**"[A-Z]['‘]\*[a-zé§¢ﬂﬁ]+['‘]\*[a-zé§¢ﬂﬁ]\*[-][A-Z][a-zé§¢ﬂﬁ]+[']\*[a-zé§¢ﬂﬁ]\*[,.:]\*\s[A-Z]\*\s\*[A-Za-z01][,.:]+\s"**)

pattern2\_1=re.compile(**" [A-Z][a-z]+"**)

## Author family name with a prefix

Texts: "McCutcheon. G: B."

Solution: Include a new case of regular expressions

pattern1\_2=re.compile(**"[A-Z]['‘]\*[a-zé§¢****ﬂﬁ]+['‘]\*[a-zé§¢ﬂﬁ]\*[A-Z][a-zé§¢ﬂﬁ]+[']\*[a-zé§¢ﬂﬁ]\*[,.:]\*\s[A-Z]\*\s\*[A-Za-z01][,.:]+\s"**)

pattern2\_1=re.compile(**" [A-Z][a-z]+"**)

## Author name as a pseud.

Texts: " Sa'rllpfr, pseud."

Solution: Include a new case of regular expressions

pattern1\_3=re.compile(**"[A-Z]['‘]\*[a-zé§¢ﬂﬁ]+['‘]\*[a-zé§¢ﬂﬁ]\*[,.:]\*\spseud."**)

## Author name with “Sir”

Texts: "McCutcheon. G: B."

Solution: Include a new case of regular expressions

pattern1\_4 = re.compile(**"[A-Z]['‘]\*[a-zé§¢ﬂﬁ]+'\*[a-zé§¢ﬂﬁ]\*[,.:]\*\sSir\s[A-Za-z01][,.:]+\s"**)

pattern2\_2 = re.compile(**"[.:]\s[A-Z][a-z]+"**)

## Author family name with two words

Texts: "Du Puy, W: A."

Solution: Include a new case of regular expressions

pattern1\_5=re.compile(**"[A-Z]['‘]\*[a-zé§¢ﬂﬁ]+['‘]\*[a-zé§¢ﬂﬁ]\*\s[A-Z][a-zé§¢ﬂ]+['‘]\*[a-zé§¢]\*[,.:]\*\s[A-Z]\*\s\*[A-Za-z01][,.:]+\s"**)

## Author name with abbreviations recognized as number:

Texts: Kerliqdall, 0.; Hannah. 1. C.

Solution: Slightly change the main pattern of regular expressions

pattern1 = re.compile(**"[A-Z]['‘]\*[a-zé§¢ﬂﬁ]+['‘]\*[a-zé§¢ﬂﬁ]\*[,.:]\*\s[A-Z]\*\s\*[A-Za-z01][,.:]+\s"**)

## Author name with a “ ' \‘ ”:

Texts: Altihe'ier, J. A.; Kelli‘dailé 0.

Solution: Slightly change the main pattern of regular expressions

pattern1 = re.compile(**"[A-Z]['‘]\*[a-zé§ﬂﬁ]+['‘]\*[a-zé§]\*[,.:]\*\s[A-Z]\*\s\*[A-Za-z01][,.:]+\s"**)

## Author name where the “.” in the first initial is missed

Texts: Hewlett. M H:

Solution: Slightly change the main pattern of regular expressions

pattern1 = re.compile(**"[A-Z]['‘]\*[a-zé§ﬂﬁ]+['‘]\*[a-zé§]\*[,.:]\*\s[A-Z]\*\s\*[A-Za-z01][,.:]+\s"**)

# Type of errors remained:

## Lines in disorder:

Ex1:

Texts: 'Casey, P. and T', "Hyde. (Ag '16"

Result: These two lines were omitted

Note: The title of this book was wrongly assigned to somewhere else and cannot be found. Therefore, although "Casey, P." is recognized by the program as "author", no titles are found, so it is omitted

Ex2:

Texts: 'Hollow tree nights and days.'

Result: This line was omitted.

Note: Author is missed.

## Cannot confirm as an author (not a recognizable pattern)

Ex1:

Texts: "eacocke, I. M. M"; 'ngehouse, P. G. Pamith '

Result: This line was omitted.

Note: First character is not characterized

Ex2:

Texts: ‘W1 n.6 K. D’; 'Row1)and, H: G.'; "(F Hag-)is, C. M."; "Rollsmeg.) S., pseud."

Result: These lines were omitted.

Note: Numbers and punctuations are included and not recognizable

Ex3:

Texts: "Singmaster. Emmeline.", 'Harcourt, .'; "Powys, .T: C."; "Forman, J M."; "Mikels. R . R."; "Sologub, F.Ea"

Result: These lines were omitted.

Note: The author’s name does not include regular pattern of initials following family names, nor sign of pseudonym name.

Ex4:

Texts: "Ha ard, H: R."; "Fits , G: "

Result: These lines were omitted

Note: One character is recognized as space.

Ex5:

Texts: "MasoniBG. S., and Hilliard, J: N. Golden hops."; "KuprinI A. I."

Result: This line was omitted

Note: Characters in family name mistakenly recognized as characterized.

## Cannot confirm as a title (not a recognizable pattern)

Ex1:

Texts: "Benson, E: F: ichael\_ (F '17)"

Result: This line was omitted.

Note: Although author is recognizable, the title is no capitalized letter in the beginning.

## Parenthesis are missed

Texts: 'Comstock H. T. Vindication. A ‘16', "Coxon. Individual. (Ap '16)( p )"

Result: The second line is recognized as a part of title of the first line (Author: “Comstock H. T.”; Title: “Vindication. A ’16 Coxon. Individual.”; Time: “(Ap '16)( p )”).