Methodological considerations

In the Genette's perspective, the title is a « threshold ». Interpreting this threshold (threshold to what?) asks for reading the novel, so that the relation between the title and the novel to be correctly assessed. With several hundreds, or a thousand novels, this becomes virtually impossible to do, at least for a single person. Still, it can be done by a group of persons – provided everybody agrees about a set of descriptors to characterize the relation between the title and the contents. This would be much more of a "crowd reading" approach (Idmhand et al, 2019), than a distant reading, but could maybe be achieved in ELTEC. In this case, the main goal of the paper would be to present and justify an annotation schema for titles.

The other way to go is a "pure" distant reading approach, that looks at intrinsic features of titles in Paul Nulty's way: length and POS composition. While we are working with much smaller collections than his[1], the added value of our study would be, in this case, in the focus on a specific genre, the novel, and in the comparison of titles from different European countries. Also, to the difference to Nulty who keeps to a very descriptive approach, we could sketch some interpretations on the basis of the observed trends.

The two approaches are not necessarily exclusive one of another!

Titles, as thresholds, are phrases characterized by a tension between precision and interest. On the one hand, a title is a summary, in a metonymic relation with the contents; while this relation can be more or less playful, it needs to be, at least to a certain extent, appropriate, to offer a description of the contents. But this description needs, in the meantime, to catch the attention of a reader, and moreover of a buyer. Therefore, a title reveals assumptions about what an author thinks is of interest for his contemporaries. The question here is to identify, using both manual annotation and maybe some automated analysis, the patterns indicating how titles describe their contents, and what is perceived of interest by writers at certain periods.

With regards to the descriptive aspect, we should probably dig deeper the morpho-syntactic composition of titles. One of the main findings in Nulty's blog, following Moretti's idea, is that titles tend to become shorter as the number of published books increases. To a certain extent, this is confirmed by our dataset, in which titles length seldom exceeds 7 words, with a mean of 5 words/ title. To me, this is related to the fact that, on the whole, titles tend to become less descriptive, or rather less verbose in the descriptions. But, as descriptions become shorter, do they become more efficient? And what are the grammatical tools used to become so?

To assess the "interest-rousing function" of titles, we could maybe take a "sentiment analysis" approach. Indeed, in some cases, reader's interest is obviously thought to be stirred by the semantic connotations of the words employed. When "diamonds" or "crimes" are mentioned in titles, it is obvious that the authors count on catching our attention by the mention of the exceptional. The fact is that I have never experimented with a sentiment analysis tool and have no idea how to use it; also, I am unsure about dictionaries for sentiment analysis being available for all the languages we cover. Maybe we could at least study the role of the adjectives as a proxy for assessing the attention of the reader with connotated aspects.

---

[1] Actually, the set we've put together from 6 languages is some 770 times smaller than his. Admittedly, we look at a much shorter period (80 years instead of 200 years), and at titles from a specific genre (the novel), but the difference remains significative, even when compared to Moretti's dataset (7000 titles).

The other approach we have taken is the annotation of titles. There is still much to say, probably, about our titles based on the annotations we have performed in the first place, but actually I felt the need to brew some tags myself, because the first ones did not answer to questions I've thought of interest, such as how often characters are mentioned in a title (as opposed to other types), are they single or groups, identified by proper names or not, etc.

After a first go to the French collection, I came out with the following tags:

CHAR: indicates if the title refers to an entity that can be considered a character.

PLA: indicates if the title refers to a place

ACT: indicates if the title mentions an action ("Le crime de l'omnibus")

PER: indicates if the title mentions a period in time, either historical or mundane ("Les plaisirs et les jours")

OTHER: indicates if the title refers to other entities than a character, a place, a period or an act (ex. "La nouvelle espérance")

In some ways, this annotation is redundant to the one we have performed in the first round, since place and characters were already indicated. But it is more fine grained and as I perform it in a single column called "Reference", it allows to see combinations. Therefore, each title is described with one or more annotations, such as in "Dodo: A Detail of the Day" that becomes "CHAR, PER". The total number of annotations (A=422) is higher than the total number of titles (T=321) by some 30%.

NB: combinations have not been studied in what follows, because I have to get back to the annotations to decide in what order to put the tags (alphabetical or as they occur in title?), and to make sure this rule is evenly respected.

In addition, I have added a column "Number", meant to be tagged with SG, PL or SO. "SG" indicated that the titles refers to a single character. "PL" is for two characters or groups. "SO" ("sans objet", in French) is when there is no character in title.

The column "NP" ("nom propre") concerns characters again. It is more or less the same thing as "person entity", but receives a "YES" only when there is a proper name to designate the character. The other tags are "NO" and again "SO".

Another new column is "Narrator", but I must say I am a bit deceived by this. With a very few exceptions such as "Mon village", the narrator appears only when there is a descriptive subtitle ("roman", "novel", etc.). I also considered a narratorial voice manifested itself when titles where double ("X or Y"). But finally this column is the most redundant with the first round of annotation.

Last, I have duplicated the column "person status" because it contained interesting information about the characters.

The new annotation has been performed in the same excel files (FRA, ENG, ROM, POR, ITA, SPA), but on a new added sheet. Also, there is an empty column between the columns I've added and the former ones.

Observations on the entire dataset

Putting all titles together, we arrive at a total of 321 items. They are quite unevenly distributed, with the T2 sub-collection being the best represented, and the T1 the least represented. This was already visible in the Malaga slides, but not stated as such. This is to be borne in mind when reading what follows, since certain trends may be significantly affected by an increase in the composition of the collections.

Repartition of titles in the 6 ELTEC collections

■ T1 ■ T2 ■ T3 ■ T4
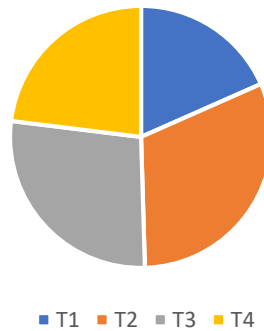
Regardless to the period, characters occupy the predominant place in titles, being present in almost three quarters of cases (70,09%). Places and acts appear much less (19% and 10,28% respectively), and periods of time even lesser (9,66%). The interesting position is that of "other", that appears in 22,43% of cases[2].

Characters in titles are in most part of cases alone (73,33%). The favorite way of designating them is the proper name (37,38% of cases), followed closely by a social status (36,14%). The two can be combined, as we will see it better when looking closely at certain collections. Taking into account the fact that masculine genre seems predominant, as shown by the data gathered in the first round of annotation, one can draw the conclusion that the "standard" (or the favorite?) way of entitling a novel in the large second half of the 19th century is the mention in the title of a eponym character, probably the main character, and most often than not a male character. This should be further confronted with some histories of the world literature. From a French point of view, it is a bit surprising, since we are told that the "roman personnel" (the novel of a person) declines after 1830. While our data is not sufficient to contradict this affirmation, it calls to nuance it, showing that a good story is still that of an individual even in the second part of the 19th century.

As for a narratorial voice, it appears in a bit more than a third of the cases (34,27%).

More interesting than these global descriptions is the evolution one can observe in the titling style over the four ELTEC periods. Indeed, while trends are to be taken with a pinch of salt, considering the dimensions and the unbalanced character of our collection, one can however notice that, on the whole, the place of the character in title tends to decrease, while the "other" category tends to become more and more visible. Another element of interest is the diminishing of places and acts in the titles of the third and the fourth subperiod, as opposed to the first two sub-periods. Could this be related to the lowering of the interest, after 1880, for a certain "sensational novel", that will play on the interest for extraordinary acts (crimes, adventures, travels), in various places, sometimes exotic (Arkansas, Africa, the desert…)?

---

[2] It is normal that percentages do no add up to 100%, since the calculations are made dividing the number of annotations (ex. CHAR=225) by the total number of titles (T=321).

Titles composition by period
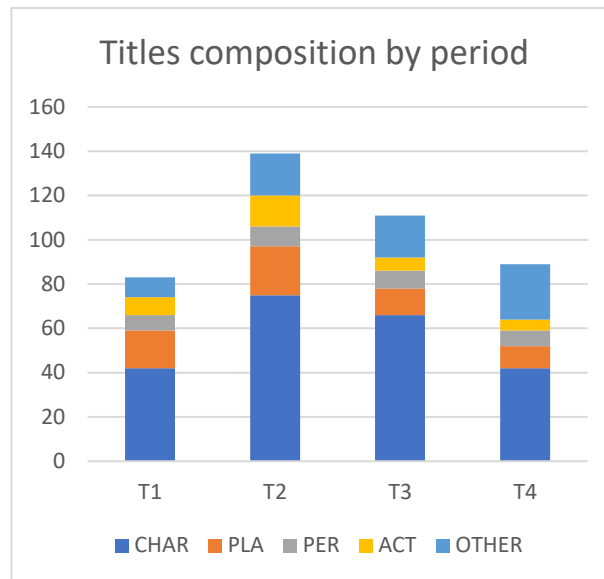
Not of a lesser interest is the share of PER in these graphs, that tends to appear quite stable. Is this an indication about the place of the historical novel, that would appear to be a constant sub-genre in ELTEC collection?

Rather than (badly) answering these questions, I'll concentrate in what follows on three of the largest collections, looking at the same set of features. Indeed, it makes no sense to look closer to the smaller collections (ROM, ITA and SPA) since a split on sub-periods leads, in these cases, to such a small number of titles, that makes any statistic senseless.

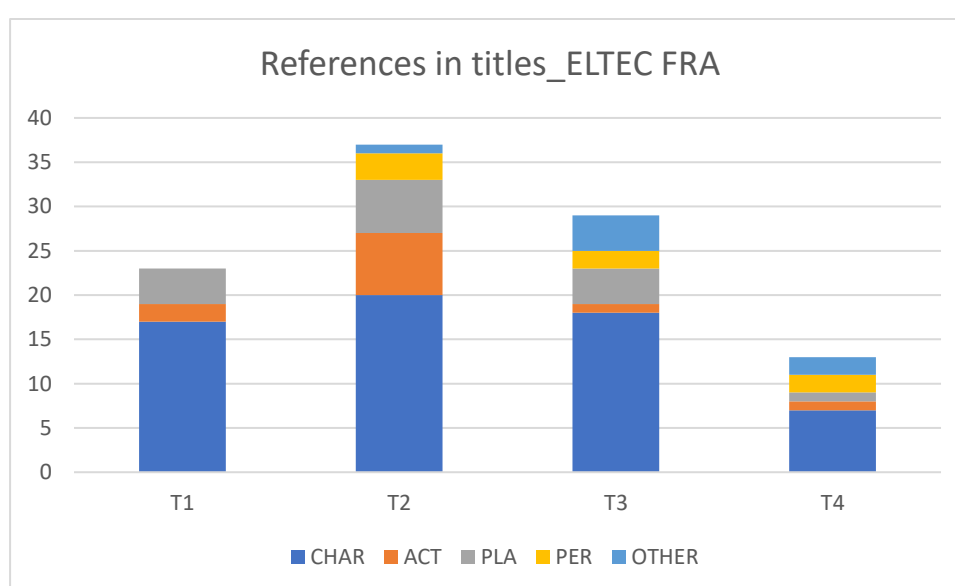**Observations on the French corpus**

As the entire collection, the French one is somewhat unbalanced, with T2 and T3 periods much better represented than the first and the last, as it can be seen in the following figure.

Number of titles per time period in ELTeC-French
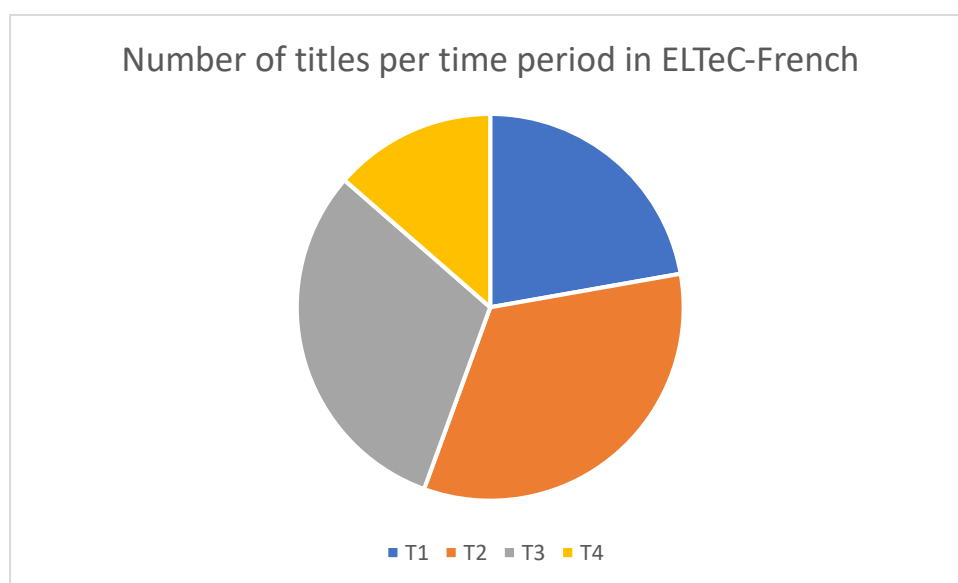
Moreover, the figures we are dealing with are quite small (between 11 and 27 novels per period), and therefore the percentages given in what follows are to be taken with much care, since they could vary as the collection grows.

In spite of these caveats, the French collection appears to illustrate in a very interesting way the above-mentioned trends. The overwhelming presence of the character tends to diminish over time, and new practices in titles (i.e., "PER" and "OTHER" tags) tend to appear more and more. This can be correlated with the fact that proper names decrease in titles (from 58,82% in T1 to 42,86% in titles), and that single characters become, in T4, almost as numerous as the groups of characters (4 SG and 3 PL tags in T4, as opposed to 15 SG and 2 PL tags in T1).

To the difference of the global dataset, the French collection indicates also a curious decrease of the number of places mentioned in titles, that has to be further documented and understood.
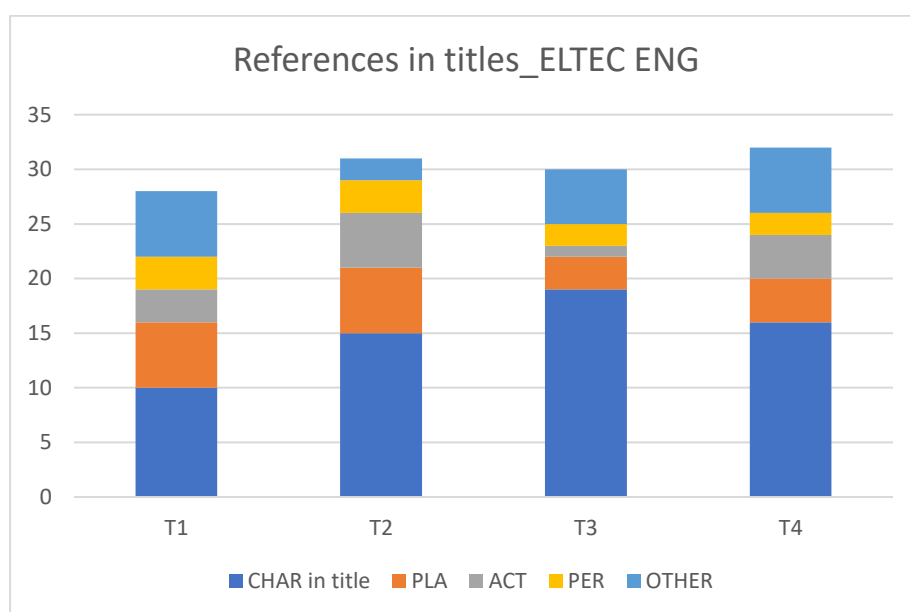

Observations on the English collection

The English collection is better balanced than the French one, in spite of a slight over-representation of T4.
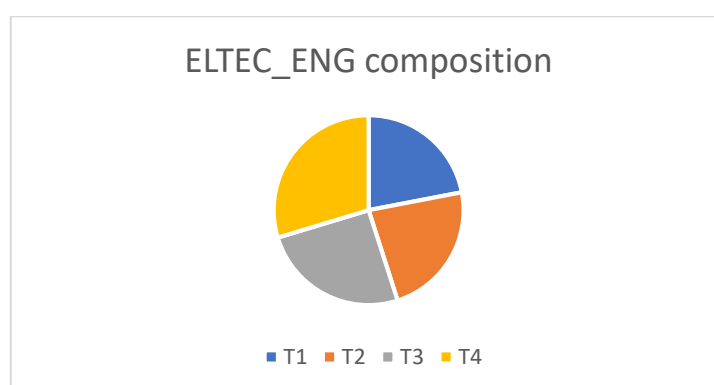
On the one hand, this collection confirms the importance of the characters in titles, as well as the fact that these characters will be rather singular ones and designated with a proper name. On the other hand, the practices seem to follow a very different curve than the French ones, with the "OTHER" tag being quite present in the first subperiod, becoming less used in the second one, and then making quite a comeback from 1880 on. Also, characters are less numerous in the first period and increase in the subsequent ones, with a spike in T3 when titling practices tend to resemble to T2 titling practices in France. Put in a very blunt and debatable way, the cross reading of these two collections suggests thus that English authors tend first to be influenced by French authors, then vice-versa. To be confirmed, infirmed, or nuanced by readings of literary histories and a closer look to data!

References in titles_ELTEC ENG

(Stacked bar chart showing values for T1, T2, T3, T4 with series: CHAR in title, PLA, ACT, PER, OTHER)

Observations on the Portuguese collection

The Portuguese collection is, again, a quite unbalanced one, with T1 and T3 under-represented.

Composition of ELTEC-POR collection

(Pie chart with segments: T1, T2, T3, T4)
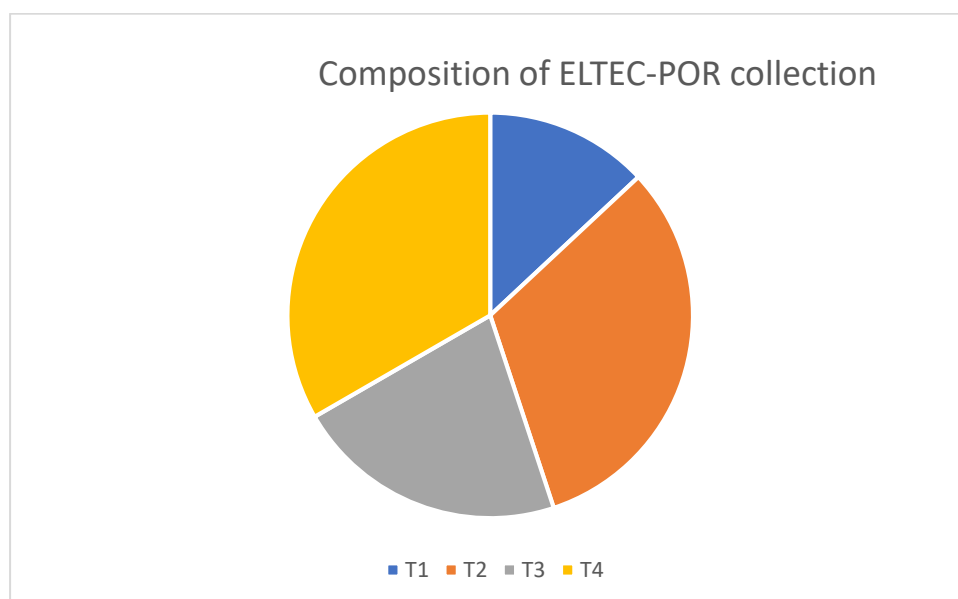
Titling styles seem, on the one hand, closer to the English than to the French ones, with the character gaining in importance in T2-T4 as opposed to T1. On the other hand, the OTHER category, inexistent in T1, is making a spectacular appearance in T2 and increases in T4 – which suggests an evolution closer to the French novel. A specific element can be the importance PER gains over time. Finally, the Portuguese collection confirms that ACT tends to disappear in titles as the 19th century unfolds.

References in tiles_ELTEC-POR