



**HAL**  
open science

# Referential Saliency in French and Mandarin Chinese: Influence of Syntactic, Semantic and Textual Factors

Jiaqi Hou, Frédéric Landragin

► **To cite this version:**

Jiaqi Hou, Frédéric Landragin. Referential Saliency in French and Mandarin Chinese: Influence of Syntactic, Semantic and Textual Factors. *Languages*, 2024, 9 (2), pp.1-26. 10.3390/languages9020040 . halshs-04416054

**HAL Id: halshs-04416054**

**<https://shs.hal.science/halshs-04416054>**

Submitted on 25 Jan 2024


**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

## Article

# Referential Salience in French and Mandarin Chinese: Influence of Syntactic, Semantic and Textual Factors

Jiaqi Hou <sup>1,\*</sup> and Frédéric Landragin <sup>2</sup> <sup>1</sup> Language Centre, Tsinghua University, Haidian District, Beijing 100084, China<sup>2</sup> Lattice, CNRS, ENS, PSL Research University, University of Sorbonne Nouvelle Paris 3, 1 Rue Maurice Arnoux, 92120 Montrouge, France; frederic.landragin@ens.psl.eu

\* Correspondence: houjiaqi@mail.tsinghua.edu.cn

**Abstract:** In this article, we propose a multifactorial approach to salience analysis, examining the influence of five factors on the salience of referential entities in discourse. Significance tests and Cramer's V tests were conducted to analyze textual data obtained through manual annotation of four text excerpts in French and in Chinese. The results show that almost all the factors have a significant influence on referents' salience (except the animacy factor in one of the excerpts). While it seems difficult to predict a fixed ranking of salience factors, which depends more on textual characteristics than on differences between the two languages, the different values of each factor under investigation show an identical behavior in terms of the positive/negative contribution to salience. The results also suggest that some factors (syntactic function and syntactic parallelism) may have a more stable influence on referents' salience than other factors (animacy, mobility, and main character), potentially constrained by textual properties such as the main character's nature, its number of occurrences, and the possible existence of competing protagonists.

**Keywords:** salience; discourse reference; multifactorial analysis; annotation; accessibility



**Citation:** Hou, Jiaqi, and Frédéric Landragin. 2024. Referential Salience in French and Mandarin Chinese: Influence of Syntactic, Semantic and Textual Factors. *Languages* 9: 40. <https://doi.org/10.3390/languages9020040>

Academic Editor: Carita Paradis

Received: 24 August 2023

Revised: 1 December 2023

Accepted: 5 December 2023

Published: 25 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction<sup>1</sup>

Salience (also referred to as 'prominence') has recently attracted considerable attention in various linguistic fields (Schneidecker 2011). In this article, this notion is examined from a referential and discursive perspective (Landragin 2004; Chiarcos et al. 2011; Von Heusinger and Schumacher 2019), which concerns a property of entities in discourse representation and serves more particularly to describe the status of centrality of certain referents in the consciousness of the partners of the enunciation (Neveu 2011). In fact, the emergence of the term and its use in the field of discourse reference stem from studies using a cognitive approach around the eighties (Chafe 1976; Chafe 1994; Prince 1981; Yule 1981; Givón 1983; Ariel 1990), according to which the choice of referential expressions is directly linked to the memory process and the cognitive system that is the mental representation of discourse entities. Salience, as presumed by the speaker and perceived by the hearer, is thus applied to referential entities in a stretch of discourse, and can account for various linguistic phenomena related to the interpretation and production of language, such as the interpretation of anaphoric expressions and the choice of referential expressions. Both the speaker and the hearer collaborate in the processes of referential choice and referential understanding, and the degree of salience of a referent indicates to the speaker which referring expression to choose, and to the hearer how to find the relevant referent. From this perspective, the consideration of this notion is essential when dealing with the automatic generation or processing of referential expressions.

With a multifactorial approach to salience, our objective is first to verify the influence of five factors, namely syntactic function, syntactic parallelism, animacy, mobility, and main character, on the salience of referents. In addition, we aim to compare not only the

relative importance of these factors, but also the contributions of each categorical value of each factor (e.g., subject or direct object of the syntactic function factor). For this purpose, we used textual data annotated with the five factors. In fact, most of the above studies were based on psycholinguistic experiments or descriptive observations. However, it would be more interesting to have recourse to attested data and to situate each referential expression in its textual context. With textual data, we are able to analyze referents' salience in their context of realization and to take into account the influence of several factors at the same time. The interest of this study also lies in its contrastive and comparative approach. Previous research in centering theory (Kameyama 1986; Walker et al. 1998; Di Eugenio 1998, see also Section 2.1 for more discussion of the theory) has shown that the factors that determine the ranking of entities according to their salience may be universal or specific to the language being processed. If the five factors under examination are all likely to influence salience in French and in Chinese (Hou and Landragin 2019), we would like to know if the contribution of each factor is similar in the two typologically different languages. Furthermore, excerpts of same genre but with different characteristics (see Section 3.1) were chosen to investigate the relative importance of the factors in four different excerpts of parallel texts. If the salience factors and its operation are perhaps constrained by textual genre (Schnedecker 2021), another question is to find out whether the factors will have the same effects in texts (or excerpts) of the same genre. Through the statistical results of the annotation, we address the following questions in this article:

- Does each of the factors have a statistically significant influence on referents' salience?
- Is the relative importance of each factor (or the ranking of factors according to their importance) similar in each language?
- Is the relative importance of each factor always similar in texts (or excerpts) of the same genre?
- Do the different categorical values of a single factor all contribute to an increase in the degree of salience? If not, are the patterns (of positive/negative contribution) similar in each language (or excerpt)?

More specifically, we put forth the following hypotheses:

- While the strength of influence might vary, each factor will have a statistically significant influence on referents' salience.
- Given the inherent linguistic differences between the two languages, the relative importance of each factor may be different in French and Chinese.
- While salience factors may be influenced by the specific textual genre, we predict that the relative importance of each factor will remain largely consistent within texts of the same genre.
- Not all values of a single factor have a uniformly positive contribution to referential salience. Some values may enhance salience, while others may diminish it, but the patterns (of positive/negative contribution) are similar in each language.

In the following sections, we first discuss the notion of salience and our multifactorial approach in Section 2. Then, we present our corpora, annotation methodology, and statistical methods in Section 3. Sections 4–7 are devoted, respectively, to the results of the statistical tests of the syntactic (syntactic function and syntactic parallelism), semantic (animacy and mobility), and textual (main character) factors. The overall results are summarized in Section 8, followed by the discussion of the stability of the factors' contributions to referential salience and some theoretical implications in Section 9. We end the last section with a conclusion and research perspectives.

## 2. Referential Salience and Salience Factors

### 2.1. Salience: Main Characteristics, Related Theories, and Multifactorial Approach

In order to define the notion of salience (or prominence), Himmelmann and Primus (2015) and Von Heusinger and Schumacher (2019) proposed three fundamental characteristics of salience:

- (i) Relational (or singling-out): the prominent status is the result of competition among language units of the same level (e.g., syllables, referents);
- (ii) Dynamic: the prominent status may change;
- (iii) Structural attraction: prominent units are structural attractors in their domain.

According to Von Heusinger and Schumacher (2019), the relational principle results from the fact that an entity is considered salient only if it is more salient in relation to the other entities. In the process of interpreting anaphors, discourse referents (realized by referential expressions) are in competition with one another and it is the most salient entity that attracts the attention of the hearer and provides an anchor for the resolution of an anaphoric expression. In Example (1), the referents *zhǔrèn* ‘the director’ and *zhè běn shū de zuòzhě* ‘the author of this book’ are in competition. After the interpretation of the second clause, it is respectively *zhǔrèn* and *zhè běn shū de zuòzhě* that attract more attention of the hearer in (1) a and (1) b, and become the salient referents in their own context.

- (1) a. [主任]<sub>i</sub>把这本书的作者介绍给我, [Ø]<sub>i</sub>鼓励我珍惜这次难得的机会。  
 [Zhǔrèn]<sub>i</sub> bǎ zhè běn shū de zuòzhě jièshào gěi wǒ,  
 director BA this CLF book DE author introduce to 1sg  
 [Ø]<sub>i</sub> gǔlì wǒ zhēnxī zhè cì nándé de jīhuì.  
 encourage 1sg cherish this CLF rare DE opportunity  
 ‘The director introduced me to the author of this book and encouraged me to cherish this rare opportunity.’
- b. 主任把这本书的作者介绍给我, [Ø]<sub>i</sub>是一个好象刚毕业的小姑娘。  
 Zhǔrèn bǎ [zhè běn shū de zuòzhě]<sub>i</sub> jièshào gěi wǒ,  
 director BA this CLF book DE author introduce to 1sg  
 [Ø]<sub>i</sub> shì yī ge hǎoxiàng gāng biyè de xiǎo gūniang,  
 is a CLF seem just graduate DE little girl  
 ‘The director introduced me to the author of this book, who appeared to be a young lady who had just graduated.’

The second criterion emphasizes that the degree of salience of an entity may change as the discourse progresses. As a result, a referent considered salient enough to be the referent of an anaphoric expression at a particular time (or place) may lose (or maintain) its high-saliency status later, as a result of the influence of salience factors (see below). The third characteristic proposes that salient units may be more central in the process of structure building and may contribute to more operations or structures. This seems to be the corollary of the special attention attributed to the most salient entity and could explain the fact that a salient referent can be more easily retrieved by a reduced linguistic form.

In the literature, several theories close to the notion of salience share this perspective that certain entities are more salient (or central) than others in the consciousness of the speaker and the hearer, and that there is a correspondence between linguistic form and degree of salience. In centering theory (Grosz et al. 1995), the ‘centers’ of an utterance are the entities (or semantic objects) that link that utterance to others in the segment of the discourse in question. According to Grosz et al. (1995), each utterance has a set of forward-looking centers (C<sub>f</sub>) that are realized through the constituent expressions of an utterance (U). The elements of C<sub>f</sub> are ranked according to their relative salience. Moreover, each utterance other than the initial utterance contains a single backward-looking center (C<sub>b</sub>) which is to be chosen from the C<sub>f</sub> of the preceding utterance and represents the discourse entity with which the current utterance is most concerned. Various factors can influence the ranking of C<sub>f</sub> in an utterance. Most of the work in centering theory emphasizes the role of syntactic functions, and considers that the subject is more likely to contribute to a rise in the ranking. Other factors such as word order, subordination, and lexical semantics are also assumed to affect the ranking.

In accessibility theory (Ariel 1990), the choice of referential expressions (or accessibility markers) by the speaker tells us about the cognitive accessibility of the referent in the mental representation of the discourse. A speaker will use a high (or low) accessibility marker to encode a referent that is assumed to be accessible (or inaccessible) to the hearer. Four factors are considered to have a determining effect on the degree of accessibility:

- (i) Distance: The distance between the antecedent and the anaphor (relevant to subsequent mentions only);
- (ii) Competition: The number of competitors on the role of antecedent;
- (iii) Saliency: The antecedent being a salient referent, mainly whether it is a topic or a non-topic;
- (iv) Unity: The antecedent being within vs. without the same frame/world/point of view/segment or paragraph as the anaphor. (Ariel 1990, pp. 28–29)

In addition to the accessibility theory, the Givenness Hierarchy (Gundel et al. 1993) also intends to associate different uses of referential expressions in discourse with the cognitive status of referents in the mental representation of interlocutors. A hierarchy of six cognitive statuses ranging from ‘in focus’ to ‘type identifiable’ is suggested:

in focus >	activated >	familiar >	uniquely identifiable >	referential >	type identifiable
it	that/this/this N	that N	the N	indefinite this N	a N

According to Gundel et al. (1993), a cognitive status higher (or more to the left) in the hierarchy includes all the lower statuses, and not the reverse. For example, an entity in focus is necessarily activated, whereas an activated entity is not necessarily in focus. With this inclusive feature, the hierarchy can allow the use of a referring expression corresponding to the lower cognitive status for an entity of a higher status, which is different from the accessibility theory which considers that the choice of a marker corresponds to a given degree of accessibility in the accessibility scale.

In fact, if the cognitive status of referents is often analyzed through the observation of referential expressions, it should be pointed out that it seems more likely that the hearer establishes a referent in his mental representation of the discourse and that he relates subsequent references to this referent to his mental representation, rather than to the original linguistic expression in the text (Brown and Yule 1983). While the entities of discourse are virtually present in the mental representation of the interlocutors, the saliency, as a property of the entities, is neither tangible nor visible. It is thus difficult to learn, in a direct way, the degree of saliency of entities.

Most of the above-mentioned studies agree that the lexical form of an entity could reflect the saliency degree of a referent in its immediate context, especially for reduced lexical forms which represent salient referents. In our analysis, saliency is quite close to but different from the notion of accessibility. On the one hand, in accessibility theory, the emphasis is put on the one-to-one relationship between the form of an expression and the cognitive status of the entity to which the expression refers, with a more or less static view. We consider that the lexical form of an entity is only a reflection of the saliency degree of a referent in its immediate context. And this reflection of the saliency degree by the form of expressions is more complex than a one-to-one relation in an authentic text. Our view of this relationship is broadly consistent with that of Gundel et al. (1993), who argue that a referent with ‘in focus’ (saliency) cognitive status may be realized prototypically by reduced forms, or less frequently by other linguistic forms generally related to a less salient referent. Therefore, even if salient referents are not always introduced by reduced referential expressions, high saliency markers (anaphoric personal pronouns and zero pronouns) necessarily encode salient referents in their context of occurrence.

On the other hand, the degree of saliency does not depend solely on the four factors in accessibility theory. If the discourse entities are constantly updated by textual data, the characteristics of the pronoun, of the antecedent, of other elements (e.g., verbs and grammatical constructions) of the relevant sentences (or, more broadly, of a discourse segment), the inherent properties of the referent, and the relational properties between the antecedent and anaphoric expressions are all likely to influence saliency, hence the importance of a multifactorial approach to saliency analysis (Landragin 2004; Hou and Landragin 2019). In accessibility theory, only the distance factor has been measured quantitatively to demonstrate distributional differences between different accessibility markers (i.e., pronoun, demonstrative,



and definite description) and their antecedent. In our study, we will measure and compare several factors in two languages to understand their contribution to referential salience. By adopting this quantitative and contrastive method, which goes beyond the scope of accessibility theory, we aim to provide empirical evidence supporting the multifactorial nature of salience. This evidence will not only enhance our understanding of salience in cognitive terms but will also contribute to a better understanding of anaphora interpretation.

In our exploration of salience from a relational perspective, we consider that the salience of an entity is determined not only by the factors that are associated with the entity in question, but also by those that arise from the contexts of its potential competitors. This relational point of view is, however, taken into account by the centering theory (Grosz et al. 1995), which proposes a ranking of  $C_f$  according to their degree of salience. However, the centering theory focuses on local coherence and models the relationship between two consecutive utterances, whereas an utterance can be linked to another more previous utterance. As a result, this theory could not explain cases where an anaphoric expression that marks high salience is not linked to an entity realized by an expression in the preceding utterance (i.e., where an anaphora and its antecedent are not located in two consecutive utterances), as well as cases where two expressions that are markers of high salience are found in the same utterance. A focus on local coherence might also miss factors that have a more global influence, such as factors from the context of encyclopedic knowledge and general cognitive processes (e.g., factors associated with the inherent semantic properties of a referent). By extending the analysis beyond immediate linguistic elements to encompass broader discourse factors, our approach offers a more nuanced understanding of the anaphora–antecedent relationship.

In our conception of salience, there is no limit to the number of salient entities in a single utterance, but the durability of the high-salience status of two or more entities over the course of the processing of the entire utterance must be questioned, as the analysis of salience must also take into account the moment and progress of the current processing or production. An entity is salient in relation to its own context and through the properties (or factors) that belong to it. That is to say, high salience status is the result of an accumulation of factors related to (but not limited to) the properties of the antecedent and the anaphora, the properties of other elements (i.e., referential, verbal or other elements) in the sentence of the antecedent and in that of the anaphora (or, even more broadly, in a segment of discourse), the inherent properties of the referent, the relational properties between the antecedent and the anaphora, the situational context, etc. In Example (2), the salience status of referents cannot be established solely on the basis of the content of the first sentence. Instead, the whole situation constructed by the two sentences in (2) involves a set of potential factors (such as syntactic function, syntactic parallelism, or animacy), making the referents ‘Susan’ and ‘Betsy’ salient for being the referents of *elle* and *lui*, respectively.

- (2)
- |    |          |          |         |          |           |    |           |
|----|----------|----------|---------|----------|-----------|----|-----------|
| a. | [Susan]i | a        | offert  | un       | hamster   | à  | [Betsy]j. |
|    | Susan    | has      | given   | a        | hamster   | to | Betsy     |
| b. | [Elle]i  | [lui]j   | a       | rappelé  | que       |    |           |
|    | She      | her      | has     | reminded | that      |    |           |
|    | les      | hamsters | étaient | assez    | sauvages. |    |           |
|    | the      | hamsters | were    | quite    | wild.     |    |           |
- ‘a. Susan gave Betsy a hamster.  
b. She reminded her that hamsters are quite wild.’

[Cornish (2000)]

In this article, we consider salience as the property of a discourse entity to be more in the center of attention in relation to other entities, in the mental representation of the speaker and the hearer, at a specific moment, and in a specific context. The notion is characterized by its relational, dynamic, and structural attraction aspects. Moreover, the complexity of the notion requires a model that considers the salience from a multifactorial perspective. According to Landragin (2004), two dimensions of salience can be distinguished, namely factors related to the cognitive aspect, such as perceptual intentions, subject attention, memory or affect, and factors related to the physical aspect. The latter includes, on the one hand,

formal physical factors, such as salience due to particular syntactic constructions, syntactic function, and word order, and on the other hand semantic physical factors such as salience related to the thematic role or the theme (or topic) of the utterance. In line with this research, [Hou and Landragin \(2019\)](#) revisited salience factors and categorized factors into syntactic, semantic, textual, and pragmatic domains:

- (i) Syntactic factors: syntactic function, grammatical constructions with salience effect, syntactic parallelism, and syntactic hierarchy;
- (ii) Semantic factors: verb semantics (in the utterance of the antecedent or of the pronoun) and referents' semantic features;
- (iii) Textual factors: order of occurrence of the referents, recency (distance), frequency of occurrence of the referents, uniqueness, and main character;
- (iv) Pragmatic factors: pragmatic constraint and the given–new distinction.

The influence of multiple factors in salience analysis or in anaphora resolution has been observed in several languages, such as French ([Landragin 2004, 2015](#); [Schneidecker 2011](#)), English ([Chiarcos 2011](#)), Spanish ([Lozano 2016](#); [Martín-Villena and Lozano 2020](#)) for L2 Spanish learners, and English ([Quesada and Lozano 2020](#)) for L2 English Learners. In this study, we examine these phenomena in light of an original study of salience in Chinese, aiming to de-lineate the specific characteristics and underlying mechanisms that drive referential salience in this language, and especially in a contrastive approach (French/Chinese). It is in this multifactorial and contrastive approach that we analyze five salience factors in this study: syntactic function, syntactic parallelism, animacy, mobility, and main character.

## 2.2. Salience Factors under Investigation

After clarifying our approach to the notion of salience, we review the discussions in the literature on the factors analyzed in this study in order to examine if they have a statistically significant influence on referents' salience, and if the factors show similar or different effects in Chinese and in French. Five representative factors among all the factors discussed in [Hou and Landragin \(2019\)](#) were selected, since these factors were found to be influential in both languages we are analyzing, and they consistently appear across the corpus, ensuring a robust dataset for analysis. The other factors have not been annotated and examined, since annotating all the factors is very time consuming, and some factors, such as syntactic constructions with salience effect, verb semantics (of implicit causality), the concrete/abstract nature of referents or pragmatic constraint, have a relatively restricted occurrence or are even virtually unobservable in our quantitative analysis corpus, which proves to be quite different from the materials used in psycholinguistic studies ([Stevenson et al. 1994](#); [Sun 2014](#)). In order to analyze these factors quantitatively with a corpus-based approach, it would be better to adopt a different methodology than the one used in this research, and to consider, for example, a search of the targeted constructions in corpus databases or in a larger corpus collection built specifically for this purpose.

In the literature, it is often argued that the most salient entity in a French sentence is the one that occupies the syntactic function of the subject. This argument is put forward especially in the work on Centering Theory and confirmed by psycholinguistic experiments ([Matthews and Chodorow 1988](#); [Gordon and Chan 1995](#); [Hudson-D'Zmura and Tanenhaus 1997](#)). In these experiments, a self-paced reading test and reading comprehension test were used to show that reading time is faster when the antecedent occupies the subject function. In addition to the subject, other functions (or values of the syntactic function factor) can be ranked according to their ability to contribute positively to the salience of entities ([Grosz et al. 1995](#)).

In the above-mentioned research, direct and indirect objects are classified in the same group, and it does not distinguish between the two. According to a cognitive point of view ([Van Hoek 2007](#)), when there are two objects in the sentence, the degree of salience of the direct object (DO) and that of the indirect object (IO) differs. While the subject functions as the most salient entity (or Figure in cognitive terms) in the sentence, the DO functions as the

second most salient entity (or primary landmark in cognitive terms) and is more prominent than the other object (the secondary landmark), which yields the following hierarchy:

- (3) Subject > direct object > indirect object > other

In Chinese, the topic (if there is one in the sentence) is considered to be the function that contributes the most to a referent’s salience (Jiang 2004, 2017; Wang 2004). Although ‘topic/theme’ is primarily considered to be a pragmatic notion (Reinhart 1981) or a notion of information structure (Lambrecht 1994), and although the ‘topic–comment’ structure is universal, it should be noted that languages have different formal devices to encode it, hence the importance of distinguishing a pragmatic topic which constitutes what the comment is about in a ‘topic–comment’ structure from the syntactic topic which is the formal device of a pragmatic topic (Gundel 1988). This distinction is especially important for Chinese (Li and Thompson 1976; Huang 1992; Her 1991; Shi 2000), which is considered as a pragmatic language (Huang 1994, 2000) and a topic–prominent language (Li and Thompson 1976). This being said, a pragmatic topic is not always encoded by a syntactic topic (it can also be encoded by a syntactic subject). Syntactic topics, however, refer always to pragmatic topics. In Examples (4) and (5), the expressions *zhè kuài jiāsù de suìpiàn* (‘the accelerating fragment’) and *tā* (‘it’), which are not subjects of the sentences, constitute the syntactic topics and encode also the pragmatic topics in (4) and (5).

- (4) 对于[这块加速的碎片] topic, 舰队太空监测系统只发出了一个三级攻击警报, ...  
 Duìyú [zhè kuài jiāsù de suìpiàn] topic, jiànduì  
 as.for this CLF accelerating DE fragment fleet  
 tàikōng jiāncè xìtǒng zhǐ fāchū le yī gè  
 spatial surveillance system only issue PFV a CLF  
 sān jí gōngjī jǐngbào,...  
 three level attack alarm  
 ‘As for the accelerating fragment, the fleet’s space surveillance system issued only a level-three attack alarm, ...’  
 [Hēi’àn sēnlín ‘The Dark Forest’, Liu Cixin (excerpt)]

- (5) [它] topic [飞行的速度] subject 很慢,...  
 [Tā] topic [fēixíng de sùdù] subject hěn màn, ...  
 3SG flying DE speed very slow  
 ‘Its flying speed was very slow,...’  
 [Hēi’àn sēnlín ‘The Dark Forest’, Liu Cixin (excerpt)]

Except for the difference in the primacy of topic function in Chinese, Wang (2004) and Jiang (2004, 2017) propose the same ranking of other values as in French:

- (6) Topic > subject > object(s) > other

Another essential factor is syntactic parallelism, also called structural parallelism. This is a phenomenon whereby anaphoric pronouns prefer to co-refer to an element having the same syntactic function in the previous clause. Unlike the previous factor, which is a syntactic property of the antecedent expression, syntactic parallelism concerns both the properties of the antecedent and those of the anaphor, or more precisely a relational property between the two expressions. In the literature, this phenomenon was first observed and considered for pronouns in subject function (Grober et al. 1978; Zhu 2002), as shown in example (7), and later for the interpretation of pronouns in object function (Chambers and Smyth 1998; Jiang 2004), as shown in example (8). In our analysis, we consider that there is a parallel relationship between the antecedent and the anaphor in cases where both expressions function as subject, DO, or IO.



(7)	Jean	a	critiqué	Paul,	
	Jean	has	criticized	Paul	
	et	il	est	parti	précipitamment. (il = Jean)
	and	he	has	left	in.a.hurry
	'Jean criticized Paul, and he left in a hurry.'				
(8)	Jean	a	critiqué	Paul,	
	Jean	has	criticized	Paul	
	et	Marie	l'	a	insulté. (l' = Paul)
	and	Marie	him	has	insulted
	'Jean criticized Paul, and Marie insulted him.'				

In addition to syntactic properties, we also analyze two semantic factors, animacy and mobility, which are the inherent properties of referents. It is often discussed in the literature, particularly in cognitive linguistic and psycholinguistic approaches, that animate entities are generally more salient than inanimate entities in both French and Chinese (Lyons 1980; Comrie 1989; Langacker 1991; Pattabhiraman 1992; Hou and Sun 2005; Wang 2014). On the other hand, the semantic feature 'mobility' is less often analyzed as a salience factor. According to Talmy (2000), Zhang (2007), and Schmid (2010), movable entities are supposed to attract more attention than immovable entities and are therefore expected to be more salient. In this article, through the exploitation of corpus data, we attempt to confirm the influence of the mobility factor on salience.

In order to decide which non-human beings we consider animate, we adopted Yamamoto's (1999) criterion that animate entities must have a face. Thus, body parts of a human or an animate object will be treated as inanimate. Although body parts have a more or less animate characteristic, this animate characteristic is in fact transferred from the entire animate (or human) entity. In other words, they do not possess in themselves this animacy. For the mobility factor, Schmid (2010) and Talmy (2000) consider that immovable entities have a permanent location. In addition to this criterion, in order to distinguish movable entities from immovable ones, we consider that movable entities are those that have, undoubtedly, the ability to move, or those that undergo a change in location in our text excerpts. As shown in example (9), *tā* ('she') is considered as an animate and movable entity, while *tā de yī zhī shǒu* ('one of his hands') is considered as an inanimate and movable entity.

(9)	她蹲在他跟前，∅拉起他的一只手，∅觉得手还是热的。					
	Tā	dūn	zài	tā	gēnqián,	
	3sc	squat	at	3sc	in.front.of	
	∅	lāqǐ	tā	de	yī	zhī shǒu,
		take	3sc	DE	a	CLF hand
	∅	juéde	shǒu	háishì	rède.	
		thik	hand	still	warm	

'She crouched down in front of him, took one of his hands and saw that it was warm.'

[*Le Ventre de Paris* 'The Belly of Paris', Émile Zola (excerpt)]

The last factor analyzed—main character—is categorized as a textual factor. Sanford and Garrod (1981) consider that a particular centrality is given to main characters when interpreting anaphors in written texts. Lima and Bianco's (1999) experiments show that the textual cue of the main character is crucial for anaphoric interpretation among French students. According to their study, references to the main character are always easier to understand, irrespective of the syntactic functions of the referent. In the corpus study of Jiang (2004), it is found that when only one main character is involved in a Chinese discourse, zero anaphora may even go across clauses or sentences to refer to the main character (which is mentioned several clauses before). In our study, we determined that the main character is the most often mentioned referent in our four text excerpts.

### 3. Materials and Methods

#### 3.1. Corpus and Annotation Methodology

The corpus of this study is composed of four narrative text excerpts of relatively small size, listed in Appendix A. In these excerpts, markables were annotated manually<sup>2</sup> using the TXM software (Heiden 2010). The corpus includes both the excerpts in their original language and the corresponding translation excerpts in the other language. While the two excerpts from ‘The Belly of Paris’ (FR and CTRF) are taken from the beginning of the novel and represent typical characteristics of the narrative genre, the two excerpts from ‘The Dark Forest’ (FTRC and CH) are in the middle of a narrative science fiction novel. Thus, even though the four annotated excerpts are of same genre, they are deliberately chosen to be distinctive. A summary of the annotation information is presented in Table 1, and the factors and the annotated values of each factor are summarized in Table 2.

**Table 1.** Summary of annotation information.

Abbreviation	Excerpt	Number of Tokens	Markables <sup>3</sup>	Annotated Factors
FR (French excerpt)	<i>Le Ventre de Paris</i> ‘The Belly of Paris’	3113	High salience markers (see below) and their potential antecedents	Syntactic function, syntactic parallelism, animacy, mobility, main character
CTRF (Chinese excerpt translated from French)	<i>Bāli de dùzi</i> ‘The Belly of Paris’	3007		
FTRC (French excerpt translated from Chinese)	<i>La forêt sombre</i> ‘The Dark Forest’	2934		
CH (Chinese excerpt)	<i>Hēi’àn sēnlín</i> ‘The Dark Forest’	2685		

**Table 2.** Annotated values of each salience factor.

Factors	Values
Syntactic function	Topic, subject, DO, IO, other
Syntactic parallelism	Yes, no
Animacy	Yes (animate), no (inanimate)
Mobility	Yes (movable), no (immovable)
Main character	Yes, no

In order to annotate salience factors and to facilitate data processing, the following major preparation stages have been carried out:

- (i) Annotation of all referential expressions. Co-referential expressions are assigned the same referent identifier under the REF property, as shown in Figure 1.
- (ii) Annotation of properties for high salience markers and potential antecedents (i.e., syntactic function, animacy, and main character).
- (iii) Exporting, cleaning, and formatting text data from the TXM tool to a CSV table.
- (iv) Generating new properties (i.e., mobility<sup>4</sup> and syntactic parallelism) in the CSV table based on properties already annotated.
- (v) Data processing for statistical methods.

With respect to high salience markers, we decided to focus on two types of markers: anaphoric personal pronouns and zero anaphors. This choice is justified firstly by the fact that these markers all contain little lexical information, and, on the other hand, they are considered to be highly accessible markers (Ariel 1990) that orient the hearer towards salient referents. The zero anaphor, like its pronominal counterpart, indicates a coherence mechanism in both languages, namely that the speaker will continue to talk about a referent already salient or present in a salient situation (Kleiber 1994). This choice is also motivated

by the fact that, according to Gundel et al. (1993), the two reduced forms in a discourse inevitably encode the salient referents with the most restrictive cognitive status, even though, in a much less frequent way, other linguistic forms can also be used to realize a salient referent in narrative texts. The observation of pronominal and zero anaphors ensures that entities identified in this way must be salient in their context, so that we can perform an analysis of these entities and the factors influencing referential salience.

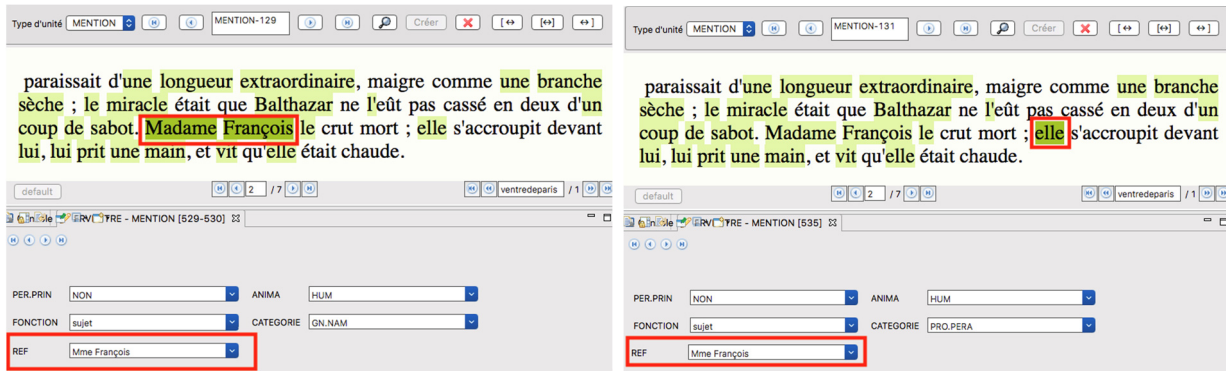


Figure 1. Coreference annotation with TXM.

Before presenting the statistical methods (Section 3.2), it is necessary to explain the influence of our approach on the data exploitation methodology. Our conception of salience is based on the fact that it is a relational notion. The high salience status of an entity exists only in comparison with other entities of the same type. If we have a series of expressions in a text (schematized by Example (10)), we consider that the referent of  $X_{n+4}$  is salient, and that this salience is determined by various factors. Analysis solely in terms of the characteristics of the anaphor ( $X_{n+4}$ ) and the antecedent ( $X_n$ ) would neglect the relational principle and the role of other potential antecedents in the process of anaphora interpretation.

- (10) Example of a sequence of referential expressions:  
 [Ding Yi] $X_n$  jeta [le marteau de [géologue] $X_{n+2}$ ] $X_{n+1}$  d'[un air abattu] $X_{n+3}$ . [Il] $X_{n+4}$  ne  
 regardait plus [la gouttelette] $X_{n+5}$ , ...  
 'Ding Yi threw the geologist's hammer with a dejected look. He no longer looked at the  
 droplet, ...'  
 Expression:  $X_n$                        $X_{n+1}$                        $X_{n+2}$                        $X_{n+3}$                        $X_{n+4}$                        $X_{n+5}$   
 Referent:     $R_n$                        $R_{n+1}$                        $R_{n+2}$                        $R_{n+3}$                        $R_n$                        $R_{n+4}$   
 Category:   NP                      NP                      NP                      NP                      Pronoun                      NP

Therefore, we examine the properties of the anaphora ( $X_{n+4}$ ), the potential antecedents ( $X_n$ ,  $X_{n+1}$ ,  $X_{n+2}$ ,  $X_{n+3}$ ), and the relationships between them ( $X_n - X_{n+4}$ ,  $X_{n+1} - X_{n+4}$ ,  $X_{n+2} - X_{n+4}$  and  $X_{n+3} - X_{n+4}$ ) in order to observe the cases of salient ( $R_n$ ) and non-salient ( $R_{n+1}$ ,  $R_{n+2}$  and  $R_{n+3}$ ) referents.

### 3.2. Statistical Methodology

In this study, the variables are of the categorical type ('salient' versus 'not salient', or the different values of each saliency factor). The Chi-squared (Chi2) test, Fisher's exact test, and Cramer's V test were applied in order to determine whether the association between the factor in question and the salience of an entity was statistically significant, and to determine the strength of this association. We also provide contingency tables and the conditional distribution of observations. In a contingency table, one variable is generally a response variable Y (the 'salience' variable in our analysis) and the other is an explanatory variable X (each salience factor). It is therefore instructive to construct a conditional probability distribution for the values of Y, given the value of X, in order to compare the various values of each salience factor.

Both the Chi2 test and Fisher's exact test aim to determine whether the two variables analyzed in a contingency table are not independent. Generally, the Chi2 test applies to large-sample data and Fisher's exact test is used when the sample size is small and especially when up to 20% of the cells have an expected number below 5. For all factors, we applied both tests in order to have a double check. The interpretation of these two significance tests is based primarily on the  $p$ -value. We chose the 0.001 significance level in order to reject null hypotheses, which are the absence of dependence between the factor in question and the salience.

Cramer's V test was used in order to measure the intensity of dependence and to make a comparison between factors, between excerpts, or between the two languages. According to Sheskin (2011), a V value below 0.3 indicates a weak association. When the V value is between 0.3 and 0.5, there is a moderate association between the two variables. And a V value greater than 0.5 indicates a high degree of dependence.

Association plots of the factors indicate the over-/under-representation of the observed frequency of a cell in a contingency table and its significance, and can help to analyze the contribution of the values taken by each factor. In an association plot, the color of the shading and the (upward or downward) orientation correspond to the (positive or negative) sign of a residual, which is used to measure the difference between the observed frequency and the expected frequency. The intensity of the shading shows its relative importance. This graph therefore makes it possible to analyze the positive/negative contribution of each value of our five factors. Multiple correspondence analysis (MCA) graphs are presented in Section 8 to visualize the relationships between salience and the five factors analyzed. Multiple correspondence analysis applies to a table which cross-classifies each individual (i.e., referential entity) with respect to all the categorical variables including the salience factors and the salience status. These MCA graphs take into account all the values (or modalities) assigned to each observation sample and represent the values often associated with a high degree of salience. We use Python 'SciPy' library and 'dython' library to calculate Chi-square tests and Cramer V-values, the R software 'vcd' library to generate association plots, and the 'Prince' library to obtain MCA graphs.

#### 4. Influence of the Syntactic Function Factor

In order to show the ranking of the various values of the factor, we first present the counts of the syntactic function versus the salience or not of the entities in Table 3. We also present the conditional distributions of the salience, given the syntactic function of the previous mention of the referent.

In both French excerpts (FR and FTRC), the subject is the function that contributes the most to increasing the referents' salience. In addition, the conditional distribution of referents realized previously by an IO shows that 50% of IOs are salient in the FR and FTRC excerpts, while the marginal percentages of salient referents are, respectively, 29.89% and 33.77%. This suggests that the IO function may contribute to referents' salience. A closer observation of the sentences containing IOs indicates that this salience may be due to the fact that an IO referent is often a human entity or even a main character in the text, at least in our four excerpts.

In the two Chinese excerpts (CTRF and CH), the topic appears to be the value that contributes the most to the increase in salience. The subject value follows closely with a conditional percentage of 56.90% of the salient antecedents in the CTRF excerpt, and 58.27% in the CH excerpt. According to the conditional distributions of the referents that are the subjects of our current investigations, the two hierarchizations of syntactic function values can be established in French (11) and Chinese (12). From the point of view of probability, a referent realized by the syntactic function further to the left of the ranking is more likely to stand out than a referent with a syntactic function value further to the right of the ranking. However, due to the relatively limited occurrence of IO and topics, a confirmatory analysis is required to enhance the reliability of the topic's and IO's positions in these rankings.

- (11) Subject > (IO >) DO > other
- (12) (Topic >) subject > (IO >) DO > other

**Table 3.** Contingency table of referents’ salience and antecedents’ syntactic function, with conditional percentages.

Excerpt	Syntactic Function	Salience		Total
		Yes	No	
FR	Subject	52.41%	47.59%	187 (100%)
	IO	50.00%	50.00%	20 (100%)
	DO	26.74%	73.26%	86 (100%)
	Other	11.72%	88.28%	239 (100%)
	Total	29.89%	70.11%	532 (100%)
CTRF	Topic	62.50%	37.50%	8 (100%)
	Subject	56.90%	43.10%	348 (100%)
	IO	40.00%	60.00%	5 (100%)
	DO	15.92%	84.08%	157 (100%)
	Other	5.29%	94.71%	227 (100%)
	Total	32.48%	67.52%	745 (100%)
FTRC	Subject	80.33%	19.67%	61 (100%)
	IO	50.00%	50.00%	2 (100%)
	DO	23.81%	76.19%	42 (100%)
	Other	13.82%	86.18%	123 (100%)
	Total	33.77%	66.23%	228 (100%)
CH	Topic	66.67%	33.33%	6 (100%)
	Subject	58.27%	41.73%	139 (100%)
	DO	13.56%	86.44%	59 (100%)
	Other	8.13%	91.87%	123 (100%)
	Total	31.50%	68.50%	327 (100%)

In order to test whether the influence of the syntactic function factor is significant and to determine the degree of intensity of this influence, we then performed the Chi2 test, the Fisher’s exact test, and the Cramer’s V test. The results in Table 4 suggest that, for all four excerpts, the dependence between the salience of a referent and the syntactic function of the antecedent is significant ( $p < 0.001$ ). The Cramer’s V values of the four text excerpts are between 0.38 and 0.58. Applying Sheskin’s (2011) criteria, the influence of the syntactic function factor on salience can therefore be classified as moderate (the FR excerpt) or strong (the CTRF, FTRC, and CH excerpts).

**Table 4.** Results of Chi2, Fisher’s exact, and Cramer’s V tests for the factor syntactic function.

Excerpt	<i>p</i> -Value (Chi2)	<i>p</i> -Value (Fisher)	Cramer’s V
FR	<0.001	<0.001	0.38
CTRF	<0.001	<0.001	0.52
FTRC	<0.001	<0.001	0.58
CH	<0.001	<0.001	0.53

Graphically, this association can be seen in the association plots (Figure 2). In the four plots, the use of enhanced shading for the bars representing the subject and other functions demonstrates that these two values all contribute significantly to the association between syntactic function and salience: subject antecedents are significantly more frequent and other antecedents are significantly less frequent in the salient group than in the non-salient group. With respect to the rest of the functions, the bars for each function have the same orientation (up or down) in all four excerpts: while there is an over-representation<sup>5</sup> of topics and IOs in the salient antecedents’ group, there is an under-representation of DOs in the salient group.



That being said, the topic and IO functions seem to be able to contribute to increasing the referents' salience, while the DO function decreases the salience degree.

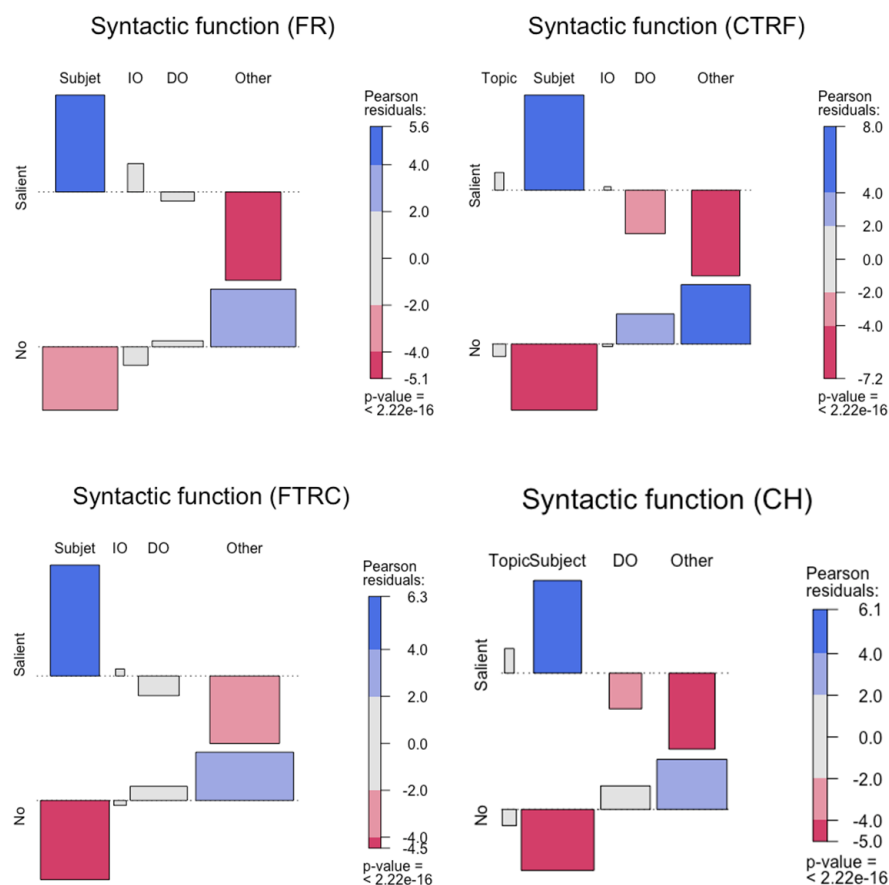


Figure 2. Association plots for the factor syntactic function.

### 5. Influence of the Syntactic Parallelism Factor

After having cross-classified the syntactic functions of potential antecedents and the referents' salience status, we seek in this section to observe the influence of another syntactic factor, namely syntactic parallelism. Unlike the factor syntactic function, syntactic parallelism is a variable that contains only two values: parallelism or not. In Table 5, we present the counts cross-classifying syntactic parallelism and salience, as well as the conditional distribution of salient and non-salient referents according to whether the anaphor and the antecedent occupy the same syntactic function or not. In the four excerpts, having syntactic parallelism is more likely to contribute to the salience of the referents, making it possible to establish the ranking of the two values in both French and Chinese:

- (13) Syntactic parallelism > no parallelism

According to the results of the Chi2 tests and Fisher's exact tests in Table 6, the influence of the syntactic parallelism factor on referents' salience is significant. Cramer's V values (respectively, 0.34, 0.49, 0.57, 0.54 in the FR, CTRF, FTRC, and CH excerpts) show that the dependence between syntactic parallelism and salience is stronger in the two excerpts of 'The Dark Forest' than in the two excerpts of 'The Belly of Paris' (the same phenomenon can be observed for the syntactic function factor, see Table 3), and that the strength of this dependence can be moderate or strong.

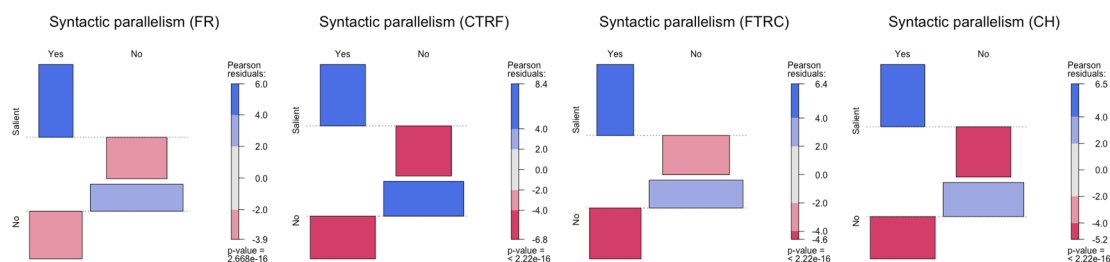
**Table 5.** Contingency table of referents’ salience and antecedents’ syntactic parallelism, with conditional percentages.

Excerpt	Syntactic Parallelism	Salience		Total
		Yes	No	
FR	Yes	58.46%	41.54%	130 (100%)
	No	20.65%	79.35%	402 (100%)
	Total	29.89%	70.11%	532 (100%)
CTRF	Yes	60.27%	39.73%	297 (100%)
	No	14.06%	85.94%	448 (100%)
	Total	32.48%	67.52%	745 (100%)
FTRC	Yes	84.91%	15.09%	53 (100%)
	No	18.29%	81.71%	175 (100%)
	Total	33.77%	66.23%	228 (100%)
CH	Yes	63.57%	36.43%	129 (100%)
	No	10.61%	89.39%	198 (100%)
	Total	31.50%	68.50%	327 (100%)

**Table 6.** Results of Chi2, Fisher’s exact, and Cramer’s V tests for the factor syntactic parallelism.

Excerpt	p-Value (Chi2)	p-Value (Fisher)	Cramer’s V
FR	<0.001	<0.001	0.34
CTRF	<0.001	<0.001	0.49
FTRC	<0.001	<0.001	0.57
CH	<0.001	<0.001	0.54

The association plots also illustrate the influence of syntactic parallelism on referents’ salience. Figure 3 shows a significant over-representation of syntactic parallelism phenomena and a significant under-representation of cases where there is no parallel relationship between anaphors and their antecedents in all four text excerpts.



**Figure 3.** Association plots for the factor syntactic parallelism.

### 6. Influence of the Semantic Features of the Referent

In this section, we step out of the syntactic domain and examine whether inherent properties of referents, such as their animate/inanimate and movable/immovable features, can influence their salience.

Tables 7 and 8 show that in all four excerpts, the proportion of salient antecedents is greater among animate entities than among inanimate entities, and the same pattern can be observed among movable and immovable entities. In that respect, we can establish the rankings of the salience degree ‘animate entities > inanimate entities’ and ‘movable entities > immovable entities’. However, are the animate (or movable) entities significantly more prominent than the inanimate (or immovable) entities in all four excerpts?

**Table 7.** Contingency table of referents’ salience and referents’ animacy, with conditional percentages.

Excerpt	Animacy	Salience		Total
		Yes	No	
FR	Yes	62.34%	37.66%	239 (100%)
	No	3.41%	96.59%	293 (100%)
	Total	29.89%	70.11%	532 (100%)
CTRF	Yes	66.03%	33.97%	315 (100%)
	No	7.91%	92.09%	430 (100%)
	Total	32.48%	67.52%	745 (100%)
FTRC	Yes	63.89%	36.11%	36 (100%)
	No	28.13%	71.88%	192 (100%)
	Total	33.77%	66.23%	228 (100%)
CH	Yes	45.00%	55.00%	40 (100%)
	No	29.62%	70.38%	287 (100%)
	Total	31.50%	68.50%	327 (100%)

**Table 8.** Contingency table of referents’ salience and referents’ mobility, with conditional percentages.

Excerpt	Mobility	Salience		Total
		Yes	No	
FR	Yes	49.84%	50.16%	315 (100%)
	No	0.92%	99.08%	217 (100%)
	Total	29.89%	70.11%	532 (100%)
CTRF	Yes	52.97%	47.03%	421 (100%)
	No	5.86%	94.14%	324 (100%)
	Total	32.48%	67.52%	745 (100%)
FTRC	Yes	54.24%	45.76%	118 (100%)
	No	11.82%	88.18%	110 (100%)
	Total	33.77%	66.23%	228 (100%)
CH	Yes	49.12%	50.88%	171 (100%)
	No	12.18%	87.82%	156 (100%)
	Total	31.50%	68.50%	327 (100%)

For the animacy factor, there are significantly more animate entities among the salient antecedents ( $p < 0.001$ , Table 9) in the FR, CTRF, and FTRC excerpts. In other words, animacy has a significant influence on referents’ salience in these three excerpts. On the other hand, in the CH excerpt, the  $p$  value (in both the Chi2 test and the Fisher’s exact test) is above the significance level (0.001), which fails to reject the independence hypothesis. Regarding the mobility factor, in all four excerpts, there are significantly more movable entities among the salient antecedents ( $p < 0.001$ , Table 10).

**Table 9.** Results of Chi2, Fisher’s exact, and Cramer’s V tests for the animacy factor.

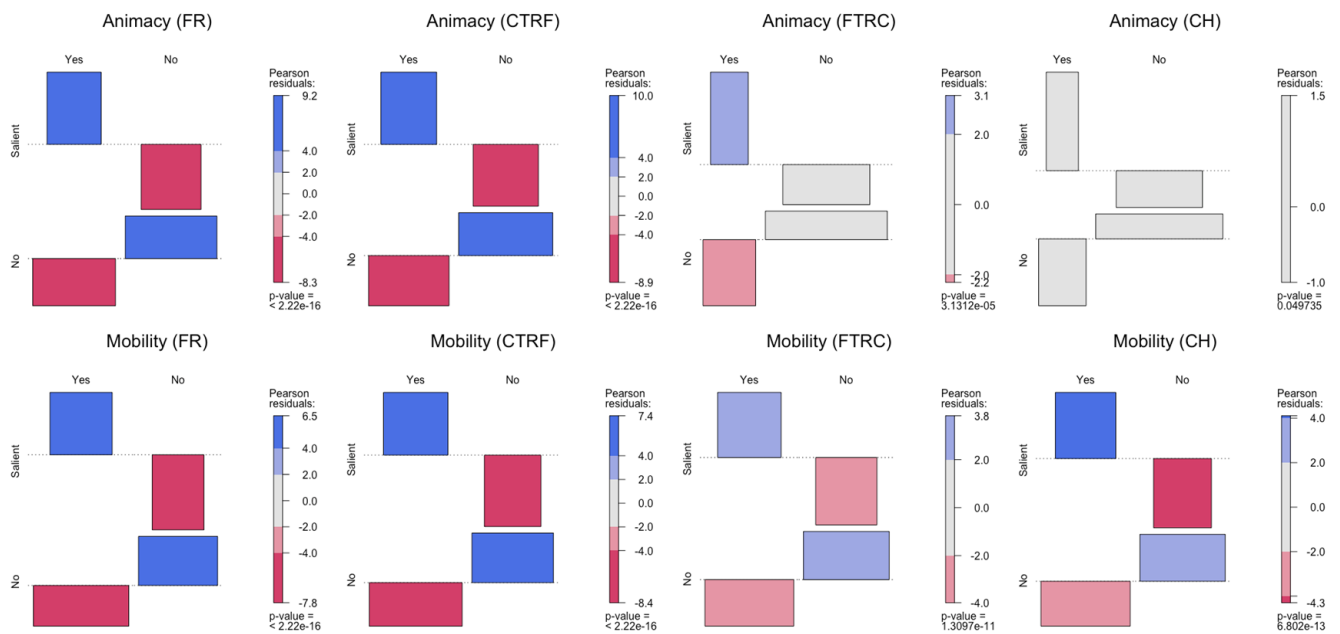
Excerpt	$p$ -Value (Chi2)	$p$ -Value (Fisher)	Cramer’s V
FR	<0.001	<0.001	0.65
CTRF	<0.001	<0.001	0.61
FTRC	<0.001	<0.001	0.25
CH	0.14	0.11	0.06

The association plots 9–16 (see Figure 4) also show that animate (movable) entities are over-represented while inanimate (immovable) entities are under-represented among salient antecedents. While the over-representation and the under-representation are significant in all four excerpts for the mobility factor and in the FR and CTRF excerpts for the animacy factor, they are not significant in the CH excerpt for the animacy factor. In the FTRC

excerpt, animate entities are significantly over-represented, but the under-representation of inanimate entities is not significant.

**Table 10.** Results of Chi2, Fisher’s exact, and Cramer’s V tests for the mobility factor.

Excerpt	<i>p</i> -Value (Chi2)	<i>p</i> -Value (Fisher)	Cramer’s V
FR	<0.001	<0.001	0.53
CTRF	<0.001	<0.001	0.50
FTRC	<0.001	<0.001	0.46
CH	<0.001	<0.001	0.42



**Figure 4.** Association plots for the factors animacy and mobility.

The Cramer’s V values in Tables 9 and 10 show that in the four excerpts, the influence of mobility on referents’ salience is more stable than that of animacy: while the association strength is between weak and strong for the animacy factor (the V values are, respectively, 0.65, 0.61, 0.25, and 0.06 in the FR, CTRF, FTRC, and CH excerpts), the degree of association is between moderate and strong for the mobility factor (the V values are, respectively, 0.53, 0.50, 0.46, and 0.42 in the FR, CTRF, FTRC, and CH excerpts).

The V values also seem to indicate that the two semantic factors play a slightly more important role in French (the FR and FTRC excerpts) than in Chinese (the CH and CTRF excerpts). Compared to the minor differences observed between the two languages, the differences are more pronounced between the two excerpts from ‘The Belly of Paris’ and the two excerpts from ‘The Dark Forest’. For both semantic features, their influence on referents’ salience is greater in the ‘The Belly of Paris’ excerpts, especially for the animacy feature. Moreover, a comparison between the V values of the two factors within the same excerpts shows that the animacy factor plays a more important role than the mobility factor in the FR and CTRF excerpts, whereas the influence of mobility is greater than that of animacy in the FTRC and CH excerpts. This could be explained by the fact that the degree of influence of the two factors may depend on the nature (semantic feature) of the main characters. While the main protagonist—‘Florent’—in the FR and CTRF excerpts is a human entity (included in the animate entity category), the main character—‘the droplet’ (a space probe)—in the FTRC and CH excerpts is a movable inanimate entity. In this context, in the latter two excerpts, there are relatively more occurrences of movable inanimate or immovable protagonists and fewer occurrences of protagonists in the upper level (in the animate category), as can be seen

in Tables 7 and 8. As a result, the influence of the animacy factor is reduced in these two excerpts, while mobility plays a more decisive role than animacy.

### 7. Influence of the Main Character Factor

In this section, we explore whether being the main character can have an influence on referents' salience. The conditional percentages in Table 11 show that in all four excerpts, the percentage of salient entities is higher when the referent is the main character than when it is another less central character (for example, 74.68% compared to 25.32% in the FR excerpt). This indicates that being the main character can promote a referent's salience.

**Table 11.** Contingency table of referents' salience and the factor main character, with conditional percentages.

Excerpt	Main Character	Salience		Total
		Yes	No	
FR	Yes	74.68%	25.32%	154 (100%)
	No	11.64%	88.36%	378 (100%)
	Total	29.89%	70.11%	532 (100%)
CTRF	Yes	72.25%	27.75%	191 (100%)
	No	18.77%	81.23%	554 (100%)
	Total	32.48%	67.52%	745 (100%)
FTRC	Yes	71.43%	28.57%	21 (100%)
	No	29.95%	70.05%	207 (100%)
	Total	33.77%	66.23%	228 (100%)
CH	Yes	67.74%	32.26%	31 (100%)
	No	27.70%	72.30%	296 (100%)
	Total	31.50%	68.50%	327 (100%)

The *p* values of the Chi2 and the Fisher's exact tests in Table 12 confirm the statistical significance (*p* < 0.001) of the influence of the main character factor. This significance is also shown in Figure 5 where there is a significant over-representation of main characters in the category of salient antecedents in all four excerpts. While less central characters are significantly under-represented in the FR and CTRF excerpts, their under-representation is not statistically significant in the FTRC and CH excerpts.

**Table 12.** Results of Chi2, Fisher's exact, and Cramer's V tests for the main character factor.

Excerpt	<i>p</i> -Value (Chi2)	<i>p</i> -Value (Fisher)	Cramer's V
FR	<0.001	<0.001	0.60
CTRF	<0.001	<0.001	0.49
FTRC	<0.001	<0.001	0.26
CH	<0.001	<0.001	0.24

With respect to the strength of association, Cramer's V values (respectively, 0.60, 0.49, 0.26, and 0.24) indicate that the association is greater in the excerpts 'The Belly of Paris'. While the effect size is rather strong in the FR and CTRF excerpts, the effect is small in the FTRC and CH excerpts. The strength of association seems to depend on the number of occurrences of the main character, since the excerpts from 'The Belly of Paris' were extracted from the beginning of the novel and contain more narration and description of the main character—'Florent'—whereas the excerpts from 'The Dark Forest' were taken from the middle of the novel and describe not only the main protagonist—'the droplet'—but also the interactions between it and the other less central protagonists. This interpretation is also supported by the number of mentions of the main character and the percentage of this number relative to the total number of mentions of referential expressions in the four excerpts, as shown in Table 13.



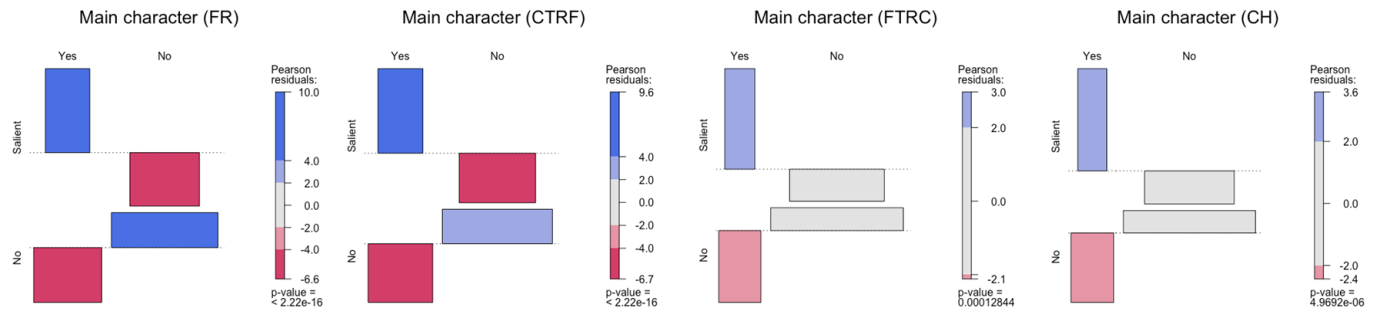


Figure 5. Association plots for the factor main character.

Table 13. Number of mentions of the main character and its percentage in relation to the total number of mentions.

Excerpt	Cramer's V	Number of Mentions of the Main Character	Percentage in Relation to the Total Number of Mentions
FR	0.60	183	22.29%
CTRF	0.49	200	22.88%
FTRC	0.26	51	7.5%
CH	0.24	52	6.75%

### 8. Overall Results and Comparison between Factors

In the previous sections, each factor was analyzed specifically and independently from the influences of the other factors. However, no single factor alone would be able to explain all the occurrences of high salience markers. In this section, we summarize the overall results and compare the contributions of the five factors in question within each text excerpt.

Firstly, we present the MCA graphs (Figure 6) of the four excerpts, which provide a synthetic visualization of the relationships between the response variable (salience) and the explanatory variables (salience factors). In the four graphs, we can see a clear opposition between referents in the high salience (salience\_YES) group and in the low salience (salience\_NO) group: on the positive side of the first factorial axis, we can notice the anaphoric expressions that represent entities of high salience; on the negative side of this axis, we see the anaphoric expressions that represent entities of low salience. The two groups (i.e., the entities with, respectively, high and low salience) are also distinguished by the over-represented values of certain factors. In the FR and CTRF excerpts, high salience is more closely related to the animate, main character, and movable values of the animacy, main character, and mobility factors (upper right corner of the plot), whereas low salience is related to the non-main character, inanimate, and immovable values (lower left corner of the plot). In the FTRC and CH excerpts, high salience is more closely associated with the subject, parallelism, and main character categories of the syntactic function, syntactic parallelism, and main character factors (lower right corner of the plot), whereas low salience is associated with the non-presence of parallelism, non-main character and other of the syntactic parallelism, and main character and syntactic function factors (upper left corner of the plot). Since component 0 (along the first factorial axis) has a greater contribution to the total inertia of the contingency table than component 1 (along the second factorial axis), all the four graphs illustrate a stronger association between high salience and subject (or topic, indirect object, syntactic parallelism, animate, movable, and main character) value, which confirms our previous analysis.

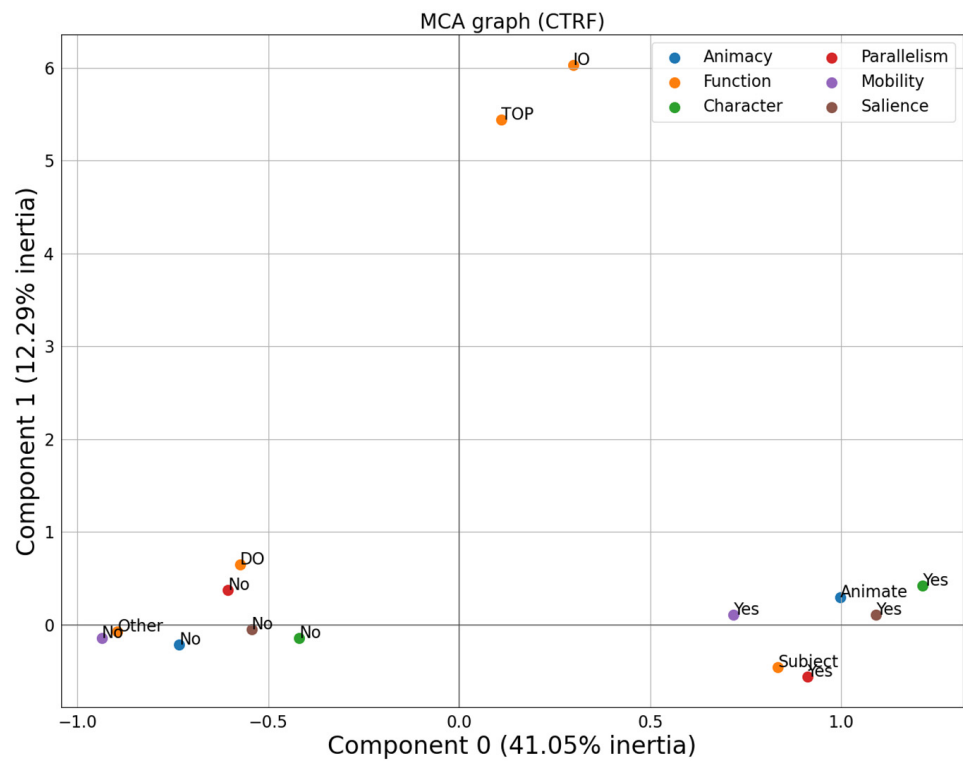
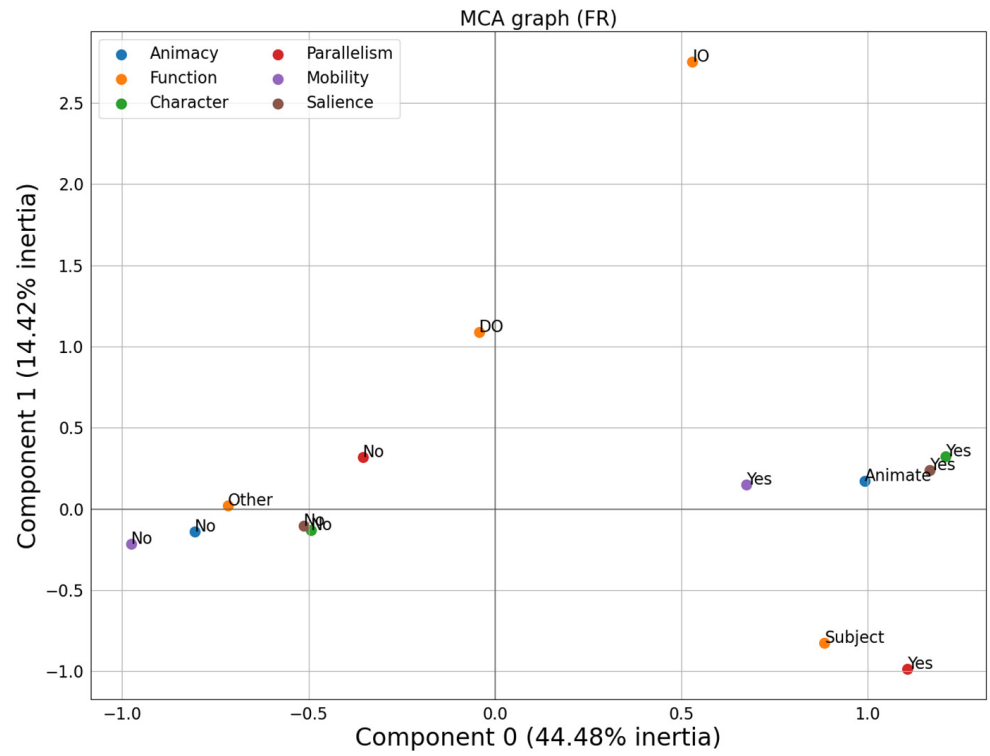
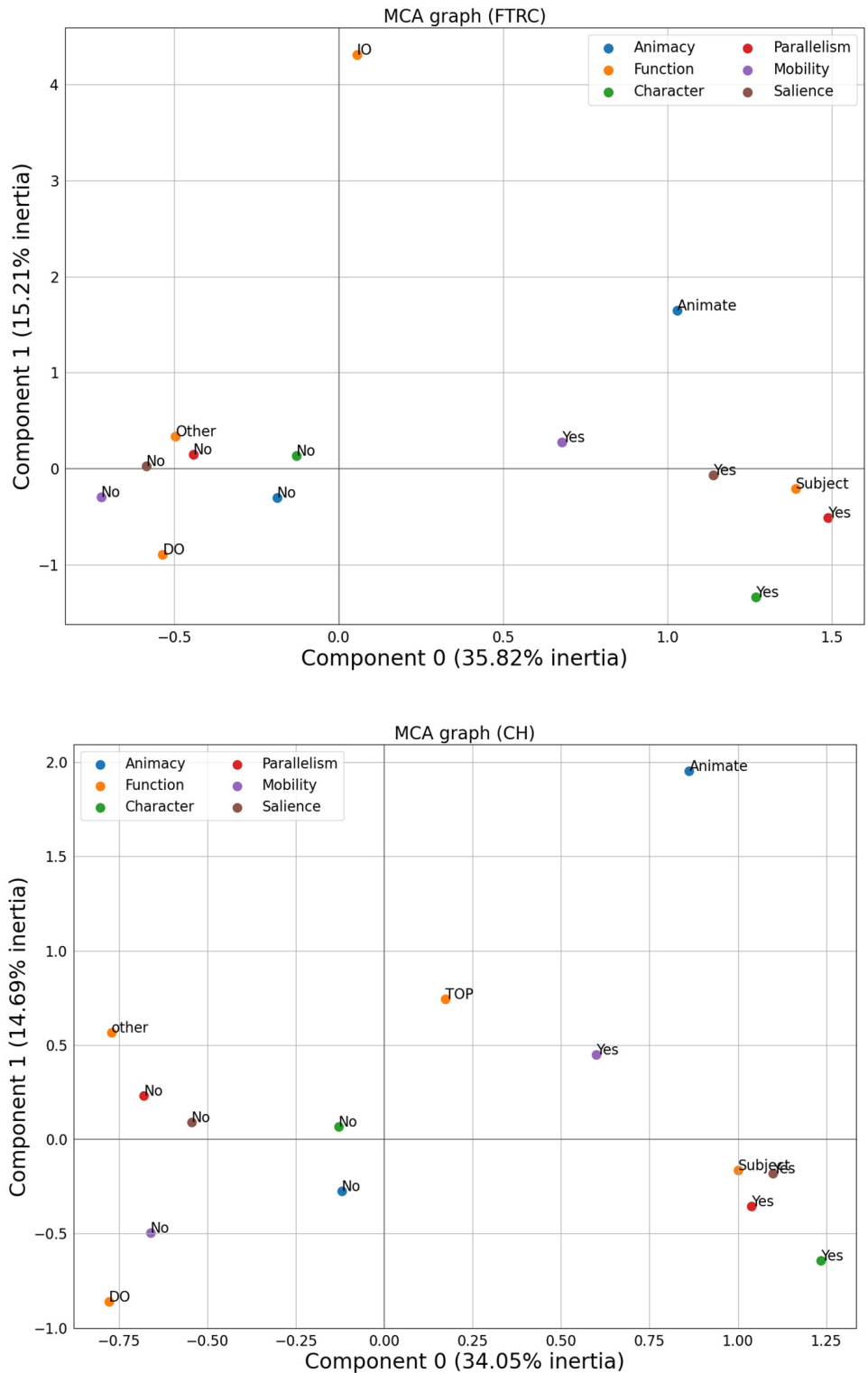


Figure 6. Cont.



**Figure 6.** Relationship between referents’ salience and factors in MCA graphs.

Table 14 summarizes the Cramer’s V values for the five factors in each excerpt. It reveals that the high salience status appears to be the result of a combination of several factors, and that this combination is not always realized in the same way: the relative importance of the factors is not always of the same order, and the relatively small effect size of one factor may be offset by an increase in the influence of other factors. For example, the small effect size of the animacy factor in the FTRC and CH excerpts could lead to the syntactic function and

syntactic parallelism factors (or some other factors that have not been analyzed in this article) playing a more important role in increasing referential salience.

**Table 14.** Summary of Cramer’s V values for all factors.

Excerpt	Factors				
	Syntactic Function	Syntactic Parallelism	Animacy	Mobility	Main Character
FR	0.38	0.34	0.65	0.53	0.60
CTRF	0.52	0.49	0.61	0.50	0.49
FTRC	0.58	0.57	0.25	0.46	0.26
CH	0.53	0.54	0.06	0.42	0.24

If we classify the factors according to V values, we obtain rather heterogeneous rankings of the factors in the four excerpts:

- (14) FR: Animacy > main character > mobility > syntactic function > syntactic parallelism
- CTRF: Animacy > syntactic function > mobility > main character > syntactic parallelism
- FTRC: Syntactic function > syntactic parallelism > mobility > main character > animacy
- CH: Syntactic parallelism > syntactic function > mobility > main character > animacy

Through the rankings of the effect sizes of the factors in (14), it seems difficult to establish a fixed ranking of salience factors, but it can be concluded that the differences due to textual characteristics (FR vs. FTRC, or CTRF vs. CH) are greater than the differences between the two languages (FR vs. CTRF, or FTRC vs. CH)<sup>6</sup>. Nevertheless, it can be noticed that Cramer’s V values for the animacy, mobility, and main character factors are slightly higher in the French excerpts than in the Chinese excerpts. Since the differences are not very pronounced, additional data will be necessary to confirm whether these factors play a more important role in French than in Chinese. While it seems impossible to predict an immutable ranking in the four excerpts, the different values of each factor under investigation show an identical behavior in terms of the positive/negative contribution to salience. In other words, we have observed a homogeneity in the rankings of values under each salience factor, as shown by the following rankings:

- (15) a Syntactic function (French): Subject > (IO >) DO > other
- a’ Syntactic function (Chinese): (Topic >) subject > (IO >) DO > other
- b Syntactic parallelism: Syntactic parallelism > non-presence of parallelism
- c Animacy: animate > inanimate
- d Mobility: movable > immovable
- e Main character: main character > non-main character

## 9. Discussion

### 9.1. Discussion of the Overall Results

In this study, we investigated the influence of various factors on referential salience in French and Chinese. Our analysis confirms that, in all the analyzed excerpts, the syntactic function, syntactic parallelism, mobility, and main character factors all have a statistically significant influence on referents’ salience. For the animacy factor, its influence is significant in most of the excerpts, except in the CH excerpt where the main character is a movable inanimate entity and animate entities have a relatively low frequency. The relative importance of each factor was not markedly different between French and Chinese. Nevertheless, it can be noticed that the animacy, mobility, and main character factors have a slightly stronger influence in the French excerpts than in the Chinese excerpts. Furthermore, we found that, even in texts of same genre, the relative importance of each salience factor can be constrained by different textual characteristics such as the nature of the main character, its number of occurrences, and the possible existence of competing protagonists. With regard to the fourth

hypothesis, our findings affirm that not all values of a single factor have a uniformly positive contribution to referential salience, but the patterns of positive and negative contributions of all values are similar in the two languages.

In addition to these findings, we would also like to discuss the stability and instability of the contributions of the factors to the salience of referents. The results in Table 14 suggest that some factors may have a more stable influence on referents' salience than other factors. On the one hand, the syntactic function and syntactic parallelism factors, whose effect sizes are between moderate and strong, contribute to the increase in salience in a reliable manner. On the other hand, a greater range (between small and strong) is found for the effect sizes of the animacy and main character factors. It is likely that the role played by these two factors, as well as the mobility factor, will vary in importance depending on the nature of the texts. As discussed in Sections 6 and 7, the influence of the animacy factor may depend on whether the main character is an animate entity, while the effect of the main character factor may be constrained by the number of times the character occurs in the excerpt in question, or by the fact that there are several competing main characters. As for the mobility factor, even if its *V* values prove to be fairly stable (between moderate and strong) in the four excerpts, it can be presumed that in a text where the main character (or rather the most central topic) is an immovable entity, the influence of the mobility factor would also be reduced, as illustrated by the description below (16) of the *Diamant dit 'le Régent'* on the website of the Louvre Museum. However, it should be noted that in narrative texts, it is not very usual to have an immovable entity as the main 'character'.

- (16) **Cette pierre** fut découverte en 1698 à Golconde, en Inde, et Ø suscita immédiatement l'intérêt de Thomas Pitt, gouverneur anglais de Madras. **Le diamant** fut taillé en Angleterre puis acquis à la demande du régent Philippe d'Orléans en 1717. **Le Régent** surpassait en beauté et en poids tous les diamants jusqu'alors connus en Occident. Aujourd'hui encore, **il** est considéré comme le plus beau diamant du monde par **sa** pureté et la qualité de **sa** taille.

'**This stone** was discovered in 1698 in Golconde, India, and Ø immediately attracted the interest of Thomas Pitt, English governor of Madras. **The diamond** was cut in England and then purchased for the French Crown at the behest of the Regent Philippe d'Orléans in 1717. **The Regent** surpassed in beauty and weight all the diamonds previously known in the western world until that time. Even today, **it** is considered to be the most beautiful diamond in the world by **its** flawless brilliance and **its** perfect cut.'

[Diamant dit 'le Régent', 'Diamond known as the "Regent"'],

<https://collections.louvre.fr/ark:/53355/cl010103121> (accessed on 9 December 2023)]

## 9.2. Theoretical Implications

In this subsection, we aim to outline some possible implications from our findings on the referential salience factors for the different theoretical frameworks in the literature. We refrain from providing a thorough analysis here, both for space reasons and because the nature of our work remains exploratory.

First of all, in a complementary approach with respect to that of the accessibility theory (Ariel 1990), we have adopted a quantitative and contrastive method and provided empirical evidence supporting the multifactorial nature of salience. The results of Chi2 and Fisher's exact tests do confirm that the salience of the entities depends on a multitude of factors, which include, but are not limited to, our five factors under investigation. Therefore, the distinction between salience and accessibility is further underscored by our findings. While accessibility theory mainly focuses on the distance, competition, saliency, and unity factors, our study reveals that salience encompasses a broader range of factors. The understanding of the influence of these factors constitutes, in fact, the reconstruction of the cues made by the speaker so that the hearer can identify the correct referent of the anaphora in question. The effectiveness of the syntactic parallelism factor also indicates that salience depends not only on factors from the characteristics of the antecedent but also on the relational properties between the antecedent and anaphoric expressions.



We then consider the implications of our findings for centering theory (Grosz et al. 1995), which considers that various factors on a rather local level (i.e., in the preceding and current utterances) can influence the salience degree. On the one hand, we have shown, through the result of animacy, mobility, and main character factors, that determining factors do not only derive from the local level, but also from a more global context that goes beyond the limit of a series of utterances, or from the context of general cognitive processes. On the other hand, our results of Cramer's V tests in Table 14 indicate that even though the five factors all contribute significantly to referential salience, their relative importance, which does not follow a fixed ranking order, depends both on textual characteristics (to a large extent) and linguistic specificities (to a lesser extent).

## 10. Conclusions and Perspectives

In this study, we have discussed the notion of salience in discourse reference, its particularities in relation to related notions such as accessibility (Ariel 1990) and centering of attention (Grosz et al. 1995), and the importance of a multifactorial analysis of salience. We also reviewed studies on the influence of five salience factors (i.e., syntactic function, syntactic parallelism, animacy, mobility, and main character), and specified for each factor our annotation criteria and annotated values.

The annotation of salience factors and the application of statistical tests (Chi<sup>2</sup>, Fisher's exact test, and Cramer's V) showed that almost all the factors have a significant influence on referents' salience (except the animacy factor in one of the excerpts). With regard to the importance of the five factors analyzed, we found that the ranking of the factors is not always of the same order and that a lower influence of one factor could be compensated by an increase in the influence of other factors. While in all four excerpts we were able to observe a regularity in the rankings of values within each salience factor, we find it difficult to predict a fixed ranking of salience factors according to their relative importance. Although our contrastive analysis of French and Chinese excerpts reveals no significant disparities in the overall importance of each factor, there are also some notable nuances to consider. For instance, the Cramer's V values for animacy, mobility, and main character factors exhibit slightly higher values in the French excerpts compared to the Chinese ones. This subtle yet important observation may offer preliminary insights into the ongoing debate regarding the language-specific factors that determine referential salience. Nevertheless, since the differences are not very pronounced, additional data will be necessary to confirm whether these factors play a more important role in French than in Chinese. Compared to the minor differences between the two languages, the importance of the factors appears to be more significantly constrained by textual characteristics such as the nature of the main character, its number of occurrences, and the possible existence of competing protagonists, at least for the five factors under investigation and in the four excerpts. For all the five factors, some categories (such as the subject category of the syntactic function factor) may enhance salience, while others (like the non-presence of parallelism of the syntactic parallelism factor) may diminish it, but the patterns (of positive/negative contribution) are similar in the two languages.

The results also indicate that certain factors (syntactic function and syntactic parallelism) may exert a more stable influence on referents' salience than other factors (animacy, mobility, and main character). The effect sizes of the latter may be constrained by textual properties such as the nature of the main character, its number of occurrences, and the possible existence of competing protagonists.

As a perspective of this work, we intend to examine, with annotation data, the influences of other salience factors (see Hou and Landragin 2019 for more discussion), such as the order of occurrence of referents in a sentence, the fact of being a pragmatic topic, and the syntactic hierarchy (i.e., main constituents versus modifiers). In addition, other methods of corpus analysis, such as corpus study using databases, can be considered to examine the effect of factors which occur less frequently in a relatively small corpus (such as the factor grammatical constructions with salience effect). An analysis of a corpus consisting of narrative texts of very different natures or texts of other genres also seems interesting to

analyze differences in terms of the importance and the stability of salience factors. This is illustrated by [Schneidecker \(2021\)](#), who points out that in informative texts of the journalistic portrait type, the main referent may rarely be taken up by a pronominal form. However, it is implausible to consider that the referent is rarely perceived as salient in readers' mental representations. In this sense, the high-salience status of referents in other textual genres would not have the same pattern of manifestation nor respond to the same factors as those observed in narrative texts. In the long term, it would be useful to explore means to capture the interactions between factors, to configure a model to classify the salience or not of referents, and thus to contribute to the interpretation of anaphoric expressions.

**Author Contributions:** Conceptualization, J.H. and F.L.; methodology, J.H.; software, J.H.; validation, J.H. and F.L.; formal analysis, J.H.; investigation, J.H.; resources, J.H.; data curation, J.H.; writing—original draft preparation, J.H.; writing—review and editing, J.H. and F.L.; visualization, J.H.; supervision, F.L.; project administration, J.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on reasonable request from the first author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

- Le Ventre de Paris 'The Belly of Paris (FR)', Émile Zola (excerpt)
- La forêt sombre 'The Dark Forest (FTRC)', Liu Cixin, translation from Chinese by Gwen-naël Gaffric (excerpt)
- Bǎlǐ de dùzi 'The Belly of Paris (CTRF)', Émile Zola, translation from French by Jīn kēnggrán and luòxuéjuān (excerpt)
- Hēi'àn sēnlín 'The Dark Forest (CH)', Liu Cixin (excerpt)

## Notes

- <sup>1</sup> Glossing abbreviations in this manuscript follow the Leipzig Glossing Rules (<http://www.eva.mpg.de/lingua/resources/glossing-rules.php> (accessed on 9 December 2023)). The following abbreviations are used in this paper: 3 = third person; CLF = classifier; DE = modification particle; DUR = durative; PFV = perfective; SG = singular.
- <sup>2</sup> The referential expressions of the excerpt *Le Ventre de Paris* 'The Belly of Paris' (FR) were annotated in the context of the DEMOCRAT project ([Landragin 2020](#)), while the other excerpts were annotated by one of the authors. In addition, all the referential expressions in the FR excerpt have been checked by the same annotator to ensure consistency and uniformity in the annotation process.
- <sup>3</sup> The markables are the linguistic units in the excerpts to which the annotations are attached.
- <sup>4</sup> The properties animacy and mobility are annotated under the same ANIMA property in the TXM annotation structure. The AN-IMA property contains four values: human entity (HUM), animate entity (ANIM), mobile inanimate entity (NON.A.M), immobile entity (NON.A.I). When generating mobility factor, we group certain values together.
- <sup>5</sup> However, it should be noted that this over-representation (or under-representation) is not always significant, probably due to a low amount of data of topic and IO functions.
- <sup>6</sup> The relatively small differences between Chinese and French could be due to the influences of the source texts on the translation texts, but it seems no less difficult to draw a more reliable conclusion with a corpus containing only original texts in French and Chinese, since it is unlikely to obtain completely comparable texts which convey the same semantic content and are pragmatically and textually equivalent, and in which all confounding factors are controlled.

## References

- Ariel, Mira. 1990. *Accessing Noun-Phrase Antecedents*. London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord. Etats-Unis d'Amérique: Routledge.
- Brown, Gillian, and George Yule. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.

- Chafe, Wallace. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics and Points of View. In *Subject and Topic*. Edited by Charles Li. New York: Academic Press, pp. 25–55.
- Chafe, Wallace. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.
- Chambers, Craig G., and Ron Smyth. 1998. Structural Parallelism and Discourse Coherence: A Test of Centering Theory. *Journal of Memory and Language* 39: 593–608. [CrossRef]
- Chiarcos, Christian, Berry Claus, and Michael Grabski. 2011. *Saliency: Multidisciplinary Perspectives on Its Function in Discourse*. Berlin and New York: Walter de Gruyter, vol. 227.
- Chiarcos, Christian. 2011. The Mental Saliency Framework: Context-adequate generation of referring expressions. In *Saliency: Multidisciplinary Perspectives on Its Function in Discourse*. Berlin and New York: Walter de Gruyter.
- Comrie, Bernard. 1989. *Language Universals and Linguistic Typology: Syntax and Morphology*. Oxford: Basil Blackwell.
- Cornish, F. 2000. L'accessibilité cognitive des référents, le centrage d'attention, et la structuration du discours: Une vue d'ensemble. *Verbum: Analecta Neolatina* 22: 7–30.
- Di Eugenio, Barbara. 1998. Centering in Italian. In *Centering Theory in Discourse*. Oxford: Clarendon Press.
- Givón, Talmy. 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language Study (Typological Studies in Language)*. Amsterdam: John Benjamins Publishing Company, vol. 3.
- Gordon, Peter C., and Davina Chan. 1995. Pronouns, passives, and discourse coherence. *Journal of Memory and Language* 34: 216. [CrossRef]
- Grober, Ellen H., William Beardsley, and Alfonso Caramazza. 1978. Parallel function strategy in pronoun assignment. *Cognition* 6: 117–33. [CrossRef] [PubMed]
- Grosz, Barbara J., Scott Weinstein, and Aravind K. Joshi. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21: 203–25.
- Gundel, Jeanette K. 1988. Universals of topic-comment structure. In *Typological Studies in Language*. Edited by Michael Hammond and Edith A. Moravcsik and Jessica Wirth Amsterdam: John Benjamins Publishing Company, vol. 17, p. 209.
- Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*. *JSTOR* 69: 274–307. [CrossRef]
- Heiden, Serge. 2010. The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme. Paper presented at the 24th Pacific Asia Conference on Language, Information and Computation, Sendai, Japan, November 4–7; pp. 389–98.
- Her, One-Soon. 1991. Topic as a grammatical function in Chinese. *Lingua* 84: 1–23. [CrossRef]
- Himmelman, Nikolaus P., and Beatrice Primus. 2015. Prominence beyond prosody: A first approximation. In *pS-prominenceS International Conference*. Viterbo: Disucom Press, University of Tuscia Viterbo, pp. 38–58.
- Hou, Jiaqi, and Frédéric Landragin. 2019. La saillance en français et en chinois: Approche multifactorielle et étude contrastive. *Linguisticae Investigationes* 42: 186–234. [CrossRef]
- Hou, Min 侯敏, and Jianjun Sun 孙建军. 2005. Zero anaphora in Chinese and how to process it in Chinese-English MT 汉语中的零形回指及其在汉英机器翻译中的处理对策. *Journal of Chinese Information Processing (中文信息学报)* 19: 15–21.
- Huang, Shuanfan. 1992. Getting to know referring expressions: Anaphor and accessibility in Mandarin Chinese. *Proceedings of ROCLING V 1992*: 27–51.
- Huang, Yan. 1994. *The Syntax and Pragmatics of Anaphora: A Study with Special Reference to Chinese*. Cambridge: Cambridge University Press.
- Huang, Yan. 2000. *Anaphora: A Cross-Linguistic Approach*. Oxford: Oxford University Press.
- Hudson-D'Zmura, Susan, and Michael K. Tanenhaus. 1997. Assigning antecedents to ambiguous pronouns: The role of the center of attention as the default assignment. In *Centering Theory in Discourse*. Oxford: Clarendon Press, pp. 199–226.
- Jiang, Ping 蒋平. 2004. Syntactic and Discourse Features of Zero Anaphora: With Specific Reference to Its Resolution in Chinese 零形回指的句法和语篇特征研究. Ph.D. thesis, Shanghai International Studies University 上海外国语大学, Shanghai, China.
- Jiang, Ping 蒋平. 2017. Accessibility Hierarchy of Antecedents in Chinese Zero Anaphora 汉语零形回指先行语的句法可及性等级序列. *Journal of Nanchang University 南昌大学学报(人文社会科学版)* 3: 135–40.
- Kameyama, Megumi. 1986. A property-sharing constraint in centering. In *Proceedings of the 24th Annual Meeting on Association for Computational Linguistics*. New York: Association for Computational Linguistics, pp. 200–6.
- Kleiber, Georges. 1994. *Anaphores et Pronoms. Louvain-la-Neuve*. Dinant: Duculot.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Landragin, Frédéric. 2004. Saillance physique et saillance cognitive. *Corela. Cognition, Représentation, Langage*, 2.
- Landragin, Frédéric. 2015. Sur Les Aspects Multicritères et Multidimensionnels de La Saillance. *Saillance. La Saillance En Langue et En Discours* 2: 15–29, Annales Littéraires de l'Université de Franche-Comté n° 940.
- Landragin, Frédéric. 2020. *Description, Modelling and Automatic Detection of Reference Chains in French*. Research Report. Paris: Agence Nationale de la Recherche.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar: Descriptive Application*. Stanford: Stanford University Press, vol. 2.
- Li, Charles N., and Sandra Thompson. 1976. Subject and Topic: A New Typology of Language. In *Subject and Topic*. Edited by Charles Li. New York: Academic Press, pp. 457–90.

- Lima, Laurent, and Maryse Bianco. 1999. Le problème des références dans la compréhension des textes à l'école primaire: Le cas de "il" et de "lui". *Revue Française de Pédagogie* 126: 83–95. [CrossRef]
- Lozano, Cristóbal. 2016. Pragmatic principles in anaphora resolution at the syntax-discourse interface. In *Spanish Learner Corpus Research: Current Trends and Future Perspectives*. Amsterdam: John Benjamins Publishing Company, vol. 78, p. 235.
- Lyons, John. 1980. *Sémantique Linguistique*. Translated by Jacques Durand, and Dominique Boulonnais. Paris: Librairie Larousse.
- Martín-Villena, Fernando, and Cristóbal Lozano. 2020. 7 Anaphora resolution in topic continuity. In *Referring in a Second Language: Studies on Reference to Person in a Multilingual World*. London: Routledge, p. 119.
- Matthews, Alison, and Martin S. Chodorow. 1988. Pronoun resolution in two-clause sentences: Effects of ambiguity, antecedent location, and depth of embedding. *Journal of Memory and Language* 27: 245–60. [CrossRef]
- Neveu, Franck. 2011. *Dictionnaire des Sciences du Langage*. Paris: A. Colin.
- Pattabhiraman, Thiagarajasarma. 1992. Aspects of Saliency in Natural Language Generation. Ph.D. thesis, Simon Fraser University, Burnaby, BC, Canada.
- Prince, Ellen. 1981. Toward a taxonomy of given-new information. In *Radical Pragmatics*. Edited by Peter Cole. New York: Academic Press, pp. 223–55.
- Quesada, Teresa, and Cristóbal Lozano. 2020. Which factors determine the choice of referential expressions in L2 English discourse? New evidence from the COREFL corpus. In *Studies in Second Language Acquisition*. Cambridge: Cambridge University Press, vol. 42, pp. 959–86.
- Reinhart, Tanya. 1981. Pragmatics and Linguistics: An analysis of Sentence Topics. *Philosophica* 27: 53–94. [CrossRef]
- Sanford, Anthony J., and Simon C. Garrod. 1981. *Understanding Written Language: Explorations of Comprehension Beyond the Sentence*. New York: Wiley.
- Schmid, Hans-Jörg. 2010. Entrenchment, Saliency, and Basic Levels. In *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford Academic. [CrossRef]
- Schneidecker, Catherine. 2011. La notion de "saliency": Problèmes définitoires et avatars. In *Saliency: Aspects Linguistiques et Communicatifs de la Mise en Évidence dans un Texte*. Olga Inkoova. Franche Comté: Presses Universitaires de Franche Comté. Available online: <https://hal.archives-ouvertes.fr/hal-00818617> (accessed on 19 April 2018).
- Schneidecker, Catherine. 2021. *Les Chaînes de Référence en Français*. Paris: Ophrys.
- Sheskin, David J. 2011. *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, London and New York: CRC Press.
- Shi, Dingxu. 2000. Topic and topic-comment constructions in Mandarin Chinese. *Language* 76: 383–408. [CrossRef]
- Stevenson, Rosemary J., Rosalind A. Crawley, and David Kleinman. 1994. Thematic roles, focus and the representation of events. *Language and Cognitive Processes* 9: 519–48. [CrossRef]
- Sun, Wen 孙雯. 2014. An ERP Study on Processing Zero Anaphor in Modern Chinese 现代汉语零形回指加工的ERP研究. Master's thesis, Jiangsu Normal University 江苏师范大学, Xuzhou, China.
- Talmy, Leonard. 2000. *Toward a Cognitive Semantics*. Cambridge: MIT Press, vol. 2.
- Van Hoek, Karen. 2007. Pronominal Anaphora. In *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford Academic. [CrossRef]
- Von Heusinger, Klaus, and Petra B. Schumacher. 2019. Discourse prominence: Definition and application. *Journal of Pragmatics* 154: 117–27. [CrossRef]
- Walker, Marilyn A., Aravind K. Joshi, and Ellen F. Prince. 1998. *Centering Theory in Discourse*. Oxford: Clarendon Press.
- Wang, Deliang 王德亮. 2004. Zero Anaphora Resolution in Chinese: A Study Based on Centering Theory 汉语零形回指解析—基于向心理论的研究. *Modern Foreign Languages* 现代外语 4: 350–59+436.
- Wang, Qian 王倩. 2014. A Study of the Cognitive Mechanism of Chinese Zero Anaphora 汉语零形回指的认知机制研究. Ph.D. thesis, Zhejiang University 浙江大学, Hangzhou, China.
- Yamamoto, Mutsumi. 1999. *Animacy and Reference: A Cognitive Approach to Corpus Linguistics*. Amsterdam: John Benjamins Publishing, vol. 46.
- Yule, George. 1981. New, current and displaced entity reference. *Lingua* 55: 41–52. [CrossRef]
- Zhang, Bojiang 张伯江. 2007. Agent and Patient: Their Semantic and Pragmatic Status and Their Roles in Chinese Grammatical Constructions 施事和受事的语义语用特征及其在句式中的实现. Ph.D. thesis, Fudan University 复旦大学, Shanghai, China.
- Zhu, Kanyu 朱勘宇. 2002. The Syntactic Motivation of Zero Anaphora in Mandarin Chinese 汉语零形回指的句法驱动力. *Chinese Language Learning* 汉语学习 4: 73–80.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.