

June 24–26, 2026

Quantitative Profiling and Cluster Analysis of Coordinators in Contemporary French

Auseichyk Yu.V.

Belarussian State University of
Foreign Languages / Minsk, Belarus
ovsei77@rambler.ru

Abstract

This study presents a quantitative methodology for constructing functional profiles of French coordinators (e.g., *et*, *mais*, *donc*, *puis*) based on their syntactic properties. Using a representative sample from the Frantext corpus (22.7 million words, 2000–present) and applying copulative (I_{cop}) and connective (I_{con}) indices, we analyse 56 units via k-means clustering. The optimal number of clusters ($k = 4$) is determined by the elbow method and silhouette score, yielding four functional zones: nuclear (4 units), near-nuclear (7), near periphery (26), and distant periphery (19), which validate a core-periphery model. Among the genuinely conjunctive uses of the units under study, the copulative and connective functions are almost equally represented; however, in the majority of contexts the units display other types of usage (adverb, discourse marker, subordinating conjunction, etc.). No direct correlation is found between frequency rank and conjunctive potential for most units outside the nuclear zone. The study offers a statistically grounded typology of French coordinators and provides a quantitative assessment of their conjunctionalisation degree.

Keywords: French coordination; Corpus analysis; Quantitative linguistics; Core-periphery model

DOI: 10.29003/2075-7182-2026-24-32-40

Количественное профилирование и кластерный анализ единиц сочинения современного французского языка

Овсейчик Ю.В.

Белорусский государственный
университет иностранных языков /
Минск, Беларусь
ovsei77@rambler.ru

Аннотация

В статье предлагается количественная методика построения функциональных профилей французских единиц сочинения (например, *et*, *mais*, *donc*, *puis*) на основе их синтаксических свойств. Материалом для исследования является репрезентативная выборка из корпуса Frantext (2000 г. – н.в., 22,7 млн словоупотреблений). С использованием копулятивного (I_{cop}) и коннективного (I_{con}) индексов и кластерного анализа (k -средних) анализируются 56 единиц, проявляющих в современном французском языке союзную функцию. Эмпирически подтверждается модель «ядро – периферия». Оптимальное число кластеров ($k = 4$), определенное на основе анализа внутрикластерной суммы квадратов (WCSS) и коэффициента силуэта, позволило выделить четыре функциональные зоны: ядерная (4 единицы), окооядерная (7 единиц), ближняя периферия (26 единиц) и дальняя периферия (19 единиц). Устанавливается, что среди собственно союзных употреблений рассматриваемых единиц копулятивная и коннективная функции представлены почти поровну; при этом в большинстве контекстов единицы демонстрируют иной способ употребления (наречие, дискурсивное слово, подчинительный союз и т. п.). Доказывается отсутствие прямой корреляции между частотным рангом и союзным потенциалом для большинства единиц, не входящих в ядро. В работе предлагается статистически обоснованная типология единиц сочинения современного французского языка и дается количественная оценка степени их конъюнкционализации.

Ключевые слова: средства сочинительной связи французского языка; корпусные методы анализа; квантитативные методы анализа; модель «ядро – периферия»

1 Introduction

In French linguistics, the classification of coordinating units has traditionally distinguished three types of connection: implicit, with conjunction substitutes, and proper coordinating conjunctions [1: 357]. Academic grammars reflect this distinction by opposing “classical” [8: 452], “properly coordinating” [22: 525], “simple” [7: 221] conjunctions – *et* ‘and’, *ni* ‘nor’, *ou* ‘or’, *mais* ‘but’, *or* ‘so’, *car* ‘because’, *donc* ‘therefore’ [5: 129] – against their analogues, termed “occasional conjunctions” [8: 452], “conjunctive adverbs” [22: 525], “functional equivalents” [7: 221] (*puis* ‘then’, *aussi* ‘also’, *alors* ‘then’, *par conséquent* ‘consequently’, etc.). These two groups are assumed to differ in syntactic behavior when connecting various types of conjuncts (see, e.g., on the syntactic nature of conjunctions and connectors in Russian [24: 633], and studies using a Supra-Corpus Database of Connectors [11, 12]).

Existing approaches to the differentiation of coordinating conjunctions and their analogues rely primarily on syntactic, morphological and semantic criteria (see the analytical review in [14]). They do not, however, offer a unified quantitative methodology for assessing conjunctive potential. The present study develops such a methodology for the quantitative interpretation of functional differentiation among French coordinators. A coordinator is defined as a unit that expresses a particular type of logical-semantic relation between two or more syntactic components: members of a single sentence (*copulative link*) or parts of a complex sentence and independent clauses (*connective link*) [2: 26–34]. For typological parallels, cf. “NP-coordination” and “event-coordination” [9: 18].

2 Review

Our approach rests on the recognition of quantitative relations as “an essential factor in language evolution and the structure of language rules” [21: 8] and on the core-periphery organisation of coordinators. The heterogeneity of syntactic linking means suggests that a limited set of units possess universal functional properties, while others acquire them only partially (on the ambivalent nature of coordination, see [13: 167]).

According to prototype theory [23], “best” representatives of a category are typically earlier forms [15: 113–114]. Thus, F. Brunot already distinguished between “inherited linking means” and newly formed conjunctive elements [4: 716]. The non-discrete boundaries within the class of coordinating conjunctions are confirmed in the work of a Norwegian romanist, who distinguishes prototypical (*et*, *ni*, *ou*, *mais*) and non-prototypical (*car*, *donc*, *or*) units [10: 37–43]. Prototypical French coordinating conjunctions have fixed initial position (hereinafter – *FIP*), allow both copulative and connective uses, can form double, repeated conjunctions (*et... et*, *et... et... et*, etc.), and do not combine with each other (see details in [20: 944]).

A core zone can be preliminarily identified as including primary coordinators – inherited from Latin *et*, *ou*, *ni*, *mais* or formed in Old French *car*, *donc*, *or*, – which have functioned as relators since the 9th–10th centuries. The periphery consists of heterogeneous units (adverbs, adverbial phrases, prepositional groups, participles, etc.) that acquired a conjunctive function similar to core coordinators during the historical development of French. E.g., *partant* functions as a participle in (1) and as a semantic-syntactic equivalent of the conjunction *donc* in (2) (all examples cited herein are drawn from Frantext [6]).

- (1) *Pô, Rhin et Rhône prenant source tous trois dans les Alpes et **partant** chacun dans une direction différente.*
‘The Po, Rhine, and Rhone rivers all originate in the Alps, each *flowing* in a different direction.’
- (2) *J’étais jeune, **partant** impulsif; je m’irritais parfois de cette immobilité.*
‘I was young, *therefore* impulsive; I was sometimes irritated by this immobility.’

3 Material and Methods

In the research a segment of the French National Corpus Frantext from 2000 to the present, with a volume of 22,720,243 word tokens [6] was used.

The inventory of 56 coordinators in contemporary French was established as follows. The initial list of candidates was compiled from three sources: (i) traditional French grammars [5, 8, 22] listing “proper coordinating conjunctions” and “conjunctive adverbs”; (ii) historical studies of French coordination [1, 10]; and (iii) lexicographic entries in [18, 19] that explicitly mention a conjunctive function. Only units occurring at least once in the corpus were retained; no automatic corpus annotations were used. Table 1 lists the units recorded in this corpus segment, ranked by absolute frequency.

1. <i>mais</i>	62 048	15. <i>plutôt</i>	5654	29. <i>du moins</i>	1135	43. <i>de fait</i>	340
2. <i>ou</i>	48 836	16. <i>de plus</i>	5318	30. <i>après tout</i>	910	44. <i>partant</i>	334
3. <i>et</i>	40 216	17. <i>ensuite</i>	4817	31. <i>tantôt</i>	748	45. <i>en outre</i>	314
4. <i>encore</i>	27 966	18. <i>en fait</i>	4627	32. <i>voire</i>	731	46. <i>par conséquent</i>	230
5. <i>aussi</i>	29 941	19. <i>d'ailleurs</i>	3982	33. <i>néanmoins</i>	687	47. <i>de même (que)</i>	224
6. <i>alors (que)</i>	22 095	20. <i>au moins</i>	3105	34. <i>en revanche</i>	677	48. <i>aussi bien que</i>	192
7. <i>puis</i>	20 725	21. <i>cependant(que)</i>	2712	35. <i>toutefois</i>	563	49. <i>à la vérité</i>	101
8. <i>ni</i>	14 086	22. <i> finalement</i>	2343	36. <i>effectivement</i>	559	50. <i>par suite</i>	57
9. <i>donc</i>	9670	23. <i>ou bien</i>	2222	37. <i>encore que</i>	545	51. <i>au reste</i>	82
10. <i>car</i>	9111	24. <i>sinon</i>	2194	38. <i>du reste</i>	486	52. <i>au surplus</i>	29
11. <i>soit</i>	8102	25. <i>c'est-à-dire</i>	1620	39. <i>c'est pourquoi</i>	473	53. <i>et/ou</i>	28
12. <i>ainsi (que)</i>	8488	26. <i>au contraire</i>	1523	40. <i>(à) savoir (que)</i>	448	54. <i>soit... ou</i>	22
13. <i>enfin</i>	8006	27. <i>en effet</i>	1589	41. <i>par contre</i>	442	55. <i>conséquemment</i>	22
14. <i>pourtant</i>	6877	28. <i>or</i>	1249	42. <i>sitôt</i>	402	56. <i>non moins que</i>	2

Table 1: Number of occurrences in the Frantext (2000 – present)

Units differ significantly in frequency: the frequency of some units (*mais*, *ou*, *et*, *aussi*, etc.) exceeds that of others (*voire*, *en revanche*, *partant*, etc.) by thousands of times. Some units (*par suite*, *au reste*, *au surplus*, *et/ou*, *conséquemment*, etc.) are recorded rarely. Hence the question: are the coordinators copulative and connective properties directly dependent on their frequency rank?

To ensure statistical comparability given the significant differences in the absolute frequency (hereinafter – *N*), a stratified representative sample was formed using a random mixing function to avoid sequences of contexts from works by the same author (details in [3: 490–491]). The sample totaled 9,642 contexts: 400 contexts for each of the eight highest-frequency units (ranks 1-8, $N > 10,000$), 200 contexts for mid-frequency units (ranks 9-29, $1000 < N \leq 10,000$), 100 contexts for low-frequency units (ranks 30-49, $100 < N \leq 1000$), and all available contexts for the least frequent units ($N \leq 100$). The selected number of contexts per unit was taken as 100 % for subsequent proportional calculations. Each context in the representative sample was analysed manually.

The proposed methodology aims to structure the coordination field synchronically by constructing a functional profile for each unit – an integral model that reflects the unit’s place and role in the system through its syntactic properties, i.e., the relative share of copulative vs. connective usage in forming coordination. Cf. the “portrait” of the Russian conjunction *ili* ‘or’ in [12].

The differentiation of coordinators is based on four criteria [20: 944] arranged hierarchically according to their diagnostic weight: the **filter** criterion (FIP), which has maximal diagnostic power and allows us to exclude contexts where the unit does not occupy the prototypical initial position of a structural link marker but instead performs another function; the **basic functional** criteria (copulative vs. connective link), which serve to calculate the quantitative indices and make it possible to objectify the conjunctionalisation degree of each unit; and the **additional** criteria (repetition of the coordinator and co-occurrence with core coordinators), which have lower diagnostic power and play an important role in interpreting transitional cases and in clarifying the status of units that exhibit close quantitative values.

4 Results and Discussion

Filter Criterion. The initial and mandatory filter was the unit’s occurrence in FIP. The diagnostic significance of this criterion lies in the fundamental difference between conjunctions and adverbs. In the distributional tradition, adverbs that act as linking units and express alternative relations (*soit*, *tantôt... tantôt*), adversative relations (*cependant*, *pourtant*, *néanmoins*, *toutefois*, *au reste*, *en revanche*, *d’ailleurs*), explicative relations (*en effet*, *c’est-à-dire*), consequence relations (*c’est pourquoi*, *aussi*,

partant, par conséquent, par suite), and conclusive relations (*enfin, ainsi*) are included among coordinators [5: 144]. However, unlike adverbs, a coordinator is not integrated into the conjunct: *Il avait tort, **cependant** il ne voulait pas l'admettre; Il avait tort, il ne voulait **cependant** pas l'admettre; Il avait tort **mais** il ne voulait pas l'admettre; *Il avait tort, il ne voulait **mais** pas l'admettre* [ibid]. All primary coordinators, except *donc*, exhibit a strict FIP. Due to its syntactic mobility, *donc* does not necessarily occupy the position between two conjuncts (*Je pense, **donc** je suis* or *Je pense, je suis **donc***, but not *Je pense, je suis*) [27: 127]. Peripheral units are also characterised by syntactic mobility.

If a unit occurs in a non-initial position, such contexts are excluded from the analysis as irrelevant for the conjunctive function. Conversely, FIP is necessary but not sufficient. E.g., in (3) *encore que* introduces a concessive clause, while in (4) it attaches a specifying component and functions as a coordinator, as evidenced by the tense-aspect forms of the verbs.

- (3) *Rien ici, pour le moment, qui puisse la soutenir, sinon la tendresse innocente de la petite Nini. **Encore que** le bébé finisse toujours par la fatiguer, et elle remet promptement l'enfant aux bonnes.*

‘Nothing here, for the moment, that could support her except the innocent tenderness of little Nini. *Although* the baby always ends up tiring her out, she promptly hands the child over to the nannies.’

- (4) *Elles étaient également adeptes du National Women's Party. **Encore que** l'une vertu a pu galvaniser l'autre et venger leur sort de fillettes exploitées.*

‘They were also adepts of the National Women's Party. *Moreover*, one virtue was able to galvanize the other and avenge their fate as exploited girls.’

Only contexts satisfying the *FIP* condition were treated as potential instances of conjunctive usage and processed further. Punctuation marks served as auxiliary cues for identifying clause boundaries. Given the specificities of French syntactic structure, punctuation was used exclusively for this purpose – to delimit clause boundaries – and never as an independent classification criterion; it was always combined with syntactic and semantic analysis. The results of this filtering are presented in Table 2. Note: *X* represents the frequency rank of the unit; *Y* represents the percentage of its occurrences in the *FIP* within the selected sample.

Units	X	Y	Units	X	Y	Units	X	Y	Units	X	Y
<i>mais</i>	1	100.0	<i>plutôt</i>	15	1.0	<i>du moins</i>	29	47.5	<i>de fait</i>	43	31.0
<i>ou</i>	2	100.0	<i>de plus</i>	16	6.0	<i>après tout</i>	30	33.0	<i>partant</i>	44	7.0
<i>et</i>	3	100.0	<i>ensuite</i>	17	38.5	<i>tantôt</i>	31	79.0	<i>en outre</i>	45	57.0
<i>encore</i>	4	7.75	<i>en fait</i>	18	28.0	<i>voire</i>	32	95.0	<i>par conséquent</i>	46	21.0
<i>aussi</i>	5	5.25	<i>d'ailleurs</i>	19	24.5	<i>néanmoins</i>	33	25.0	<i>de même (que)</i>	47	20.0
<i>alors (que)</i>	6	45.25	<i>au moins</i>	20	21.5	<i>en revanche</i>	34	43.0	<i>aussi bien que</i>	48	26.0
<i>puis</i>	7	71.25	<i>cependant (que)</i>	21	41.5	<i>toutefois</i>	35	32.0	<i>à la vérité</i>	49	23.0
<i>ni</i>	8	100.0	<i> finalement</i>	22	25.5	<i>effectivement</i>	36	25.0	<i>par suite</i>	50	8.77
<i>donc</i>	9	20.5	<i>ou bien</i>	23	45.5	<i>encore que</i>	37	9.0	<i>au reste</i>	51	17.07
<i>car</i>	10	100.0	<i>sinon</i>	24	68.5	<i>du reste</i>	38	16.0	<i>au surplus</i>	52	34.48
<i>soit</i>	11	17.5	<i>c'est-à-dire</i>	25	91.0	<i>c'est pourquoi</i>	39	82.0	<i>et/ou</i>	53	100.0
<i>ainsi (que)</i>	12	26.5	<i>au contraire</i>	26	25.5	<i>à savoir (que)</i>	40	33.0	<i>soit... ou</i>	54	100.0
<i>enfin</i>	13	19.5	<i>en effet</i>	27	20.5	<i>par contre</i>	41	36.0	<i>conséquemment</i>	55	7.09
<i>pourtant</i>	14	30.5	<i>or</i>	28	100.0	<i>sitôt</i>	42	1.0	<i>non moins que</i>	56	50.0

Table 2: Frequency of units in *FIP* in contemporary French (2000–present)

Basic Functional Criteria. For contexts that passed the *FIP* filter, two basic syntactic functions were differentiated: copulative (Criterion 1) and connective (Criterion 2) links.

Criterion 1 possesses greater diagnostic power than Criterion 2, since copulativity is a historically primary property of prototypical coordinators, indicating their original “emptied” meaning and high syntactic flexibility [17: 119]. Traditionally only *et*, *mais*, *ou*, *ni* are thought to connect homogeneous constituents without a formally expressed subject in the second conjunct (cf.: *il lit **et** écrit*, but *il dort, **car** il est malade* [16: 84]). Such sentences are transitional between simple and complex sentences

because they exhibit a dual nature: the predicates share a subject, yet their predicative base is split. However, our study shows that other units are also permissible in such contexts (examples (5), (6)), and some preliminarily assigned peripheral units can perform the copulative function (examples (7), (8)).

- (5) *Elle ne marqua aucun étonnement, **au contraire** fondit de gentillesse.*
‘She showed no astonishment, on the contrary melted with kindness.’
- (6) *Rafaele l’embrasse, **puis** la serre contre lui.*
‘Rafaele kisses her, then hugs her.’
- (7) *Il plissait des yeux dans la lumière mate **pourtant** éblouissante.*
‘He squinted in the dull yet dazzling light.’
- (8) *L’épaisse étoffe **ainsi que** sa chemise de nuit blanche constituaient un rempart suffisant pour sa vertu.*
‘The thick fabric and her white nightgown were sufficient protection for her virtue.’

Criterion 2 concerns the connection of independent clauses or parts of a complex sentence, each with its own formally expressed subject. When the subject is recoverable from the context through grammatical agreement (as in examples (5), (6) above), the context was classified as copulative. When the second conjunct has its own subject, the context was classified as connective (example (9)). This distinction avoids ambiguity and ensures data comparability.

- (9) *Le bruit de la mitraille le laissait froid. **Au contraire** cela m’excitait le sang et l’imagination.*
‘The noise of the machine-gun left him cold. On the contrary it excited my blood and imagination.’

Based on these two criteria, the Copulativity Index (I_{cop}) and the Connectivity Index (I_{con}) are calculated on a scale from 0.0 (minimum) to 1.0 (maximum). I_{cop} is numerically equal to the ratio of the number of contexts where the studied unit exhibits copulative properties to the selected number of contexts for that unit. I_{con} is calculated similarly for the connective properties of each unit. E.g., *puis* has $I_{cop} = 0.35$, $I_{con} = 0.36$.

Additional functional criteria.

Criterion 3 pertains to the unit’s ability to form double or repeated conjunctions. This criterion has medium, narrowly specialised diagnostic power and is factored into the calculation of the I_{cop} and I_{con} indices. The ability to form constructions (CO) X_1 CO X_2 CO X_3 is an optional but highly specific property of a strong coordinator. Its absence, however, is not critical. For example, the binary nature of the adversative relation X_1 *mais* X_2 does not structurally or semantically presuppose repetition of the coordinator. The analysis shows that repetition of coordinators in contemporary French is characteristic not only of core units *et*, *ou*, *ni* (which are designed to form sets of equal elements) but also of peripheral units. This property is illustrated in the examples below. For instance, two units with identical I_{cop} and I_{con} may differ in their ability to be repeated; the one that allows repetition is considered more highly conjunctionalised.

- (10) *...des malades qui ne peuvent attendre le lendemain, **donc** des malades en état d’urgence, **donc** de grands corps purulents qui contaminent tout ce qu’ils touchent.*
‘...patients who cannot wait until the next day, therefore emergency patients, therefore large suppurating bodies that contaminate everything they touch.’
- (11) *Un jour tu es amoureuse d’un autre monsieur, **puis** tu te re-sépars **puis** tu es encore amoureuse.*
‘One day you fall in love with another man, then you break up, then you fall in love again.’

- (12) *D'incessants déferlements de rumeur tantôt me font onde, tantôt me font sable, tantôt me font roc.*

‘Ceaseless torrents of rumor turn me now into a wave, now into sand, now into rock.’

Criterion 4 assesses a unit’s ability to combine with core coordinators. Though possessing weaker diagnostic power, it is crucial for identifying peripheral status, as core coordinators do not combine with one another. Postposition relative to a core coordinator typically indicates a modification of the connection. The combinatorial potential between coordinators necessitates further research procedures to elucidate the underlying semantic, pragmatic, and discursive factors governing their co-occurrences.

- (13) *La langue de mon père, c’était le yiddish. Aussi le russe.*

‘My father’s language was Yiddish. And also Russian.’

- (14) *Ils furent ensemble étonnés, et aussi soulagés...*

‘They were astonished together, and also relieved...’

- (15) *Elle était souvent faite d’argent, de crédits, de factures, mais aussi parfois du regard des autres, de quartiers nouveaux, de discussions judiciaires...*

‘She was often made of money, credit, bills, but also sometimes of the gaze of others, new neighborhoods, the legal proceedings...’

Based on manual annotation of the 9,642 contexts across all 56 units, the weighted sum of copulative uses equals 2,270 and the weighted sum of connective uses equals 2,262, giving a total of 4,533 genuinely conjunctive occurrences. Thus, copulative uses constitute 23.5 % of the entire sample, connective uses account for 23.4 %, and the remaining 53.1 % of the contexts are non-conjunctive (adverbial, discourse-marker, or subordinating). Among only genuinely conjunctive usage, the proportions are almost equal: 50.1 % copulative vs. 49.9 % connective.

Preliminary grouping and cluster analysis. The indices I_{cop} and I_{con} allow a tripartite division based on the inequality between them: with a greater share of copulative usage, i.e., $I_{cop} > I_{con}$ (e.g., *ni, et, aussi bien que, (à) savoir, et/ou*); with a greater share of connective usage, i.e., $I_{cop} < I_{con}$ (e.g., *c’est-à-dire, ou bien, mais, en revanche, cependant, par contre, toutefois, pourtant, en effet*); with relatively equal shares of copulative and connective usage, i.e., $I_{cop} \approx I_{con}$ (e.g., *puis, soit, sinon, au contraire, donc, partant, par suite, par conséquent, ainsi, enfin, aussi*). If absolute values are also taken into account, units can be manually ranged from those with maximal indices (1.0) to those with no conjunctive uses (0.0). However, manual grouping based on arbitrary cut-off points is problematic. Statistical verification is therefore required to identify objectively homogeneous functional groups.

To identify statistically homogeneous groups of 56 coordinators according to their functional profile, we applied *k*-means clustering (Python) to the I_{cop} and I_{con} indices. The optimal number of clusters (*k*) was determined by a combination of the elbow method (plotting within-cluster sum of squares, WCSS) and the silhouette score for $k = 2 \dots 10$. The elbow curve showed a sharp bend at $k = 4$: WCSS decreased from 2.921 ($k = 3$) to 1.914 ($k = 4$). The silhouette coefficient for $k = 4$ was 0.607. Lower k (2 or 3) merged functionally distinct groups (e.g., nuclear and near-nuclear zones), leading to a 52.6 % increase in intra-cluster variance when moving from $k = 4$ to $k = 3$. Higher k (5 or 6) produced clusters with very low separability (silhouette < 0.4) and contained fewer than 3 units, which is statistically unstable. Thus, $k = 4$ was chosen as the optimal balance between parsimony and granularity. The obtained four clusters are interpreted as: **nuclear zone** (4 units), **near-nuclear zone** (7 units), **near periphery** (26 units) and **distant periphery** (19 units). Visualization of the results is presented in Figure.

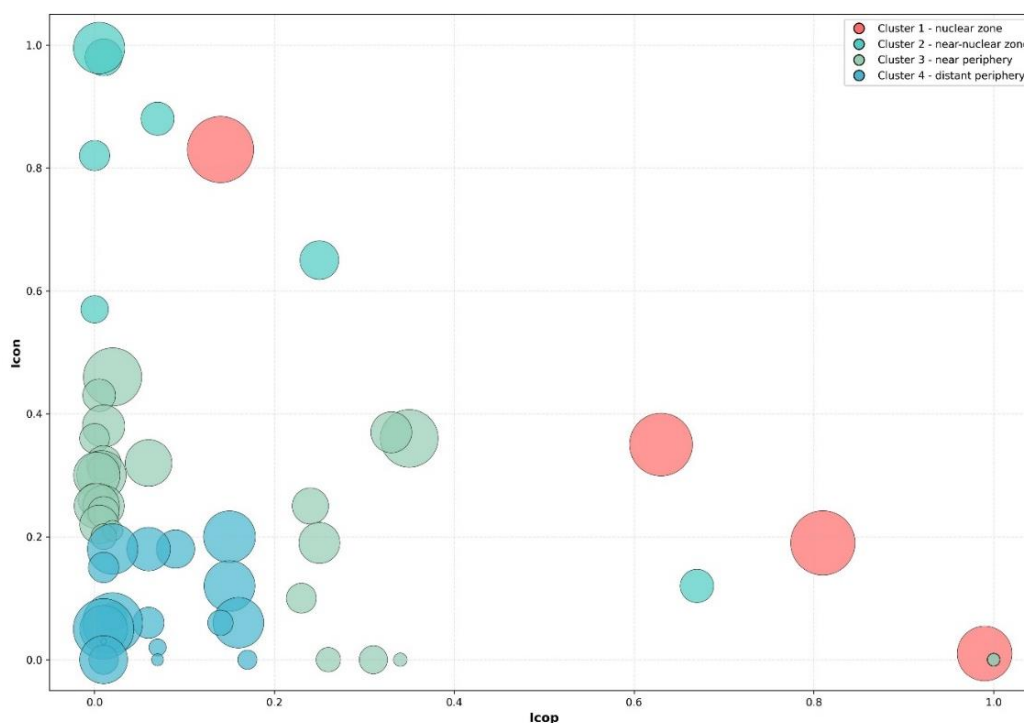


Figure: Distribution of French coordinators according to their functional profile by clusters

Marker size is proportional to the logarithm (base 10) of the absolute frequency of each unit (see Table 1). Marker colour encodes the cluster assignment obtained by *k*-means clustering. The plot demonstrates that high absolute frequency (large markers) does not automatically imply a nuclear position: frequent units such as *donc* (rank 9), *alors* (rank 6) and *puis* (rank 7) fall into peripheral clusters. Conversely, some low-frequency units (small markers) like *c'est pourquoi* (rank 39) belong to the near-nuclear zone. Thus, no direct correlation exists between frequency rank and conjunctive potential.

Nuclear Zone unites prototypical conjunctions *ni*, *ou*, *et*, *mais* with maximally high mean I_{cop} and I_{con} values. They mark basic logical-semantic relations and form diverse syntactic constructions (Criteria 1, 2, 3) and do not combine with each other (Criterion 4).

- (16) *Je ne me sentais **ni** malade, **ni** surtout la moindre envie d'abandonner.*
 'I felt neither sick nor, least of all, any desire to give up.'

Near-Nuclear Zone is formed by units with high values of one index and moderate or low values of the other, indicating functional asymmetry. They establish both types of connection (Criteria 1, 2), limitedly form (bi-) polysyndeton (Criterion 3), and usually combine with core coordinators (Criterion 4), except *or*, *car*. Three subtypes are identified: specialised connective coordinators (*or*, *car*, *c'est pourquoi*, $I_{cop} < I_{con}$); specialised copulative coordinators (*tantôt*, *c'est-à-dire*, $I_{cop} > I_{con}$); balanced coordinators (*voire*, *en outre*, $I_{cop} \approx I_{con}$).

Near Periphery comprises units exhibiting medium index values ($0.19 > k\text{-mean} < 0.46$). It represents the largest and functionally most diverse group. These units demonstrate syntactic variability (Criteria 1, 2, 3) and combine with core coordinators (Criterion 4). According to their functional profiles, two subtypes are distinguished. Specialised connective coordinators form the most numerous subgroup, characterised by a predominance of the connective function ($I_{cop} < I_{con}$) and can be further divided by the intensity of this property: units with relatively high values ($0.30 > I_{con} < 0.46$) *puis*, *sinon*, *alors*, *en revanche*, *cependant*, *ensuite*, *par contre*, *après tout*, *toutefois*, *pourtant*, *en fait*; units with medium values ($0.19 > I_{con} < 0.29$) *ou bien*, *aussi bien que*, *néanmoins*, *finalement*, *d'ailleurs*, *effectivement*, *en effet*, *à la vérité*, *par conséquent*. Specialised copulative coordinators are characterised by a high I_{cop} while the I_{con} is zero or close to zero: *et/ou*, *soit... ou*, *au surplus*, *de fait*.

Distant Periphery comprises units with values close to zero for I_{cop} and low values for I_{con} (k -mean < 0.19). Units are used occasionally as coordinator, often with stylistic or syntactic restrictions (*partant, conséquemment, non moins que*). This zone is characterized by minimal conjunctionalisation degree and includes: balanced coordinators ($I_{cop} \approx I_{con}$, *ainsi (que), au contraire, au moins, aussi, de même (que), de plus, donc, enfin, encore (que), partant, par suite, soit*); specialised coordinators with minimal functionality ($I_{cop} \approx 0$ or $I_{con} \approx 0$, *au reste, du reste, conséquemment, non moins que, sitôt, plutôt*). Notably, *donc*, initially classified as core, in contemporary French demonstrates low functional specialisation indices ($I_{cop} = 0.12$, $I_{con} = 0.19$), placing it in the distant periphery, indicating its transcategorial status.

- (17) *L'idée d'instruction, **donc** de transmission de savoirs est le cadet des soucis des responsables de l'Instruction publique.*
 'The idea of instruction, and hence of knowledge transmission, is the least of the concerns for those responsible for Public Education.'
- (18) *Pas de poubelles pour ne pas officialiser le lieu. Et si le lieu n'est pas officiel, il n'existe pas. Pas de poubelles **donc** pas de lieu.*
 'No trash cans [were placed] to avoid formalising the place. And if a place is not official, it does not exist. No trash cans, therefore no place.'

5 Conclusion

The proposed methodology provides a quantitative grounding for the functional differentiation of coordinators in contemporary French over the last two decades. By applying a hierarchical set of criteria (FIP filter, copulative vs. connective distinction, repetition and combinability) to a stratified 9,642-context sample, we have constructed functional profiles for 56 units. K -means clustering of the I_{cop} and I_{con} indices yields four robust clusters, interpreted as nuclear (4 units), near-nuclear (7), near periphery (26) and distant periphery (19). This validates the core-periphery organisation of the coordination field and allows the degree of conjunctionalisation of each unit to be quantified.

Crucially, the vast majority (53.1%) of contexts are non-conjunctive: they correspond to adverbial, discourse-marker or subordinating uses. Among the genuinely conjunctive occurrences, the copulative and connective functions are almost perfectly balanced (50.1% vs. 49.9%). Hence, there is no functional predominance of one link type over the other.

For most peripheral coordinators, the conjunctive function is not the primary mode of usage, and no direct correlation exists between frequency rank and conjunctive potential. These findings highlight the necessity of a graded, multi-criterial analysis and demonstrate that the majority of peripheral units occurrences are non-conjunctive – a result that would be entirely obscured by a simple dichotomy between copulative and connective links.

References

- [1] Antoine G. La coordination en français. — Paris : d'Artrey, 1958–1962. — 2 vol.
- [2] Auseichyk Yu.V. (2023), System of French coordinators in diachrony [Sistema edinic sochineniya francuzskogo yazyka v diahronii], Minsk, MSLU.
- [3] Baranov A.N. (2012), Linguistic examination of text: theory and practice [Lingvisticheskaya ekspertiza teksta: teoriya i praktika], Moscow.
- [4] Brunot F. La pensée et la langue. Méthodes, principes et plan d'une théorie nouvelle du langage appliquée au français. — Paris : Masson, 1956.
- [5] Dubois J., Jouannon G., Lagane R. Grammaire française. — Paris : Larousse, 2019.
- [6] Frantext. Access mode: <http://www.frantext.fr>.
- [7] Gak V.G. (2000), Theoretical grammar of the French language [Teoreticheskaya grammatika francuzskogo yazyka], Moscow: Dobrosvet.
- [8] Grevisse M., Goosse A. Le Bon Usage. Grammaire française. — 17e édition. Louvain-la-Neuve: De Boeck Supérieur, 2021.
- [9] Haspelmath M. Coordination // Language typology and linguistic description. — Cambridge, 2009. — Vol. 2 : Complex constructions. — P. 3–50.

- [10] Hobaek-Haff M. *Coordonnants et éléments coordonnés*. — Oslo : Solum Forlag ; Paris : Didier Erudition, 1987.
- [11] In'kova O.Yu. (2017), Universal classification of discourse units and supracorpus database [Universal'naya klassifikatsiya diskursivnykh edinic i nadkorpurnaya baza dannykh], Proceedings of the V.V. Vinogradov Russian Language Institute [Trudy Instituta russkogo yazyka im. V.V. Vinogradova], 12, pp. 344–367.
- [12] In'kova O.Yu. (2020), A quantitative method for analysis of connectives: the “portrait” of the russian conjunction ili (or) in the supracorpus database of connectives [Kolichestvennyy metod analiza konnektorov: «portret» russkogo soyuza ili v nadkorpurnoy baze dannykh konnektorov], Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog 2020” [Komp'yuternaya Lingvistika i Intellekturnye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii “Dialog 2020”], 19, pp. 372–386.
- [13] Kiklevich A.K. (2019), Meaning – text – interaction: linguistic representations [Smysl – tekst – interaktsiya: yazykovye reprezentatsii], Moscow: Languages of Slavic cultures [Yazyki slavyanskikh kul'tur].
- [14] Kobozeva I.M. (2025), Function words: problem areas yesterday, today and tomorrow [Sluzhebnye slova: problemnye oblasti vchera, segodnya i zavtra], Bulletin of the South Ural State University [Vestnik Yuzhno-Ural'skogo Gosudarstvennogo Universiteta], 22(2), pp. 6–17.
- [15] Kubryakova E.S. (2004), Language and knowledge. Towards acquiring knowledge of language: parts of speech from a cognitive perspective. The role of language in understanding the world [Yazyk i znanie. Na puti polucheniya znaniy o yazyke: chasti rechi s kognitivnoy tochki zreniya. Rol' yazyka v poznanii mira], Moscow: Languages of Slavic cultures [Yazyki slavyanskikh kul'tur].
- [16] Kudryavceva N.B. (2010), Controversial issues of grammar: on the composition of coordinatif conjunctions in French [Spornye voprosy grammatiki: o sostave soyuzov sochineniya vo francuzskom yazyke], Lomonosov Linguistics and Intercultural Communication Journal [Vestnik Moskovskogo universiteta, Ser. 19, Lingvistika i mezhkul'turnaya kommunikatsiya], 2, pp. 81–87.
- [17] Lambert F. Et: un cas de grammaticalisation? // Travaux de linguistique du Cerlico. — 2001. — 14. — P. 113–134.
- [18] Larousse. Access mode: <https://larousse.fr/dictionnaires/francais>.
- [19] Le Robert. Access mode: <https://www.lerobert.com/>.
- [20] Marchello-Nizia Ch. [et al.]. *Grande Grammaire Historique du Français*. — Berlin ; Boston : De Gruyter Mouton, 2020.
- [21] Plungyan V.A. (2008), The corpus as a tool and as an ideology: some lessons from modern corpus linguistics [Korpus kak instrument i kak ideologiya: o nekotorykh urokakh sovremennoy korpusnoy lingvistiki], Russian Language and Linguistic Theory [Russkiy yazyk v nauchnom osveshchenii], 2(16), pp. 7–20.
- [22] Riegel M., Pellat J.-C. , Rioul R. *Grammaire méthodique du français*. — Paris : PUF, 2016.
- [23] Rosch E. *Principles of Categorisation // Cognition and Categorization* / ed.: E. Rosch, B. Lloyd. — Hillsdale, 1978. — P. 27–48.
- [24] Uryson E.V. (2012), Conjunctions, connectors and valency theory [Soyuzy, konnektory i teoriya valentnostej] Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog 2012” [Komp'yuternaya Lingvistika i Intellekturnye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii “Dialog 2012”], 11 (18), pp. 627–638.
- [25] Wilmet M. *Grammaire rénovée du français*. — Bruxelles : De Boeck, 2007.