Research Journal of English Language and Literature (RJELAL)
A Peer Reviewed (Refereed) International Journal
Impact Factor 6.8992 (ICI) http://www.rjelal.com;
Email:editorrjelal@gmail.com; ISSN:2395-2636 (P); 2321-3108(O)

Vol.13.Issue 4. 2025 (Oct-Dec)

RESEARCH ARTICLE

INTERNATIONAL STANDARD SERIAL NUMBER INDIA
2395-2636 (Print):2321-3108 (online)

Corpus-Driven Study of Pattern Grammar in French for the Electricity Domain

Yang Tong[1], Fan Xinyan[2]
[1,2]School of Foreign Languages, North China Electric Power University, Beijing 102206, China
Corresponding Author: Yang Tong (*Email:* tongyang@ncepu.edu.cn)

DOI: 10.33329/rjelal.13.4.105

Article info
Article Received: 02/10/2025
Article Accepted: 06/11/2025
Published online: 12/11/2025

Abstract

This study adopts a corpus-driven approach to investigate the constructional properties and semantic functions of the "N à N" structure in electricity-related French. Based on a specialized corpus of technical French, high-frequency instances of this pattern were extracted using AntConc and subsequently subjected to manual contextual verification and semantic categorization. The findings reveal a strong tendency toward constructionalization and semantic stability of the "N à N" structure within specialized discourse. Beyond its canonical form, the structure frequently appears in nested and extended variants, reflecting the precision-oriented nature of technical language. Semantic analysis indicates that the "N à N" pattern predominantly encodes three major semantic domains: equipment-related, energy-related, and parameter-control categories. These domains exhibit marked register-specificity across different technical fields. The study empirically supports the form–meaning mapping central to Pattern Grammar and contributes to terminology extraction and constructional modeling in technical French.

Key words: Pattern Grammar; Corpus linguistics; Corpus driven

1. Introduction

As the founding figure of the "lexical turn" in British linguistics, J.R. Firth was the first to introduce the core concept of "meaning by collocation", which challenged the traditional semantic focus on conceptual cognition and redefined word meaning as inherently relational—emerging from patterns of co-occurrence. Within Firth's theoretical framework, collocation is not merely a lexical combination but an abstract syntactic regularity: its meaning is not derived from isolated lexical items, but is instantiated through recurrent co-occurrence patterns.

Building on Firth's theory of collocation, John Sinclair developed a systematic framework

of corpus-driven grammar through the practical implementation of the COBUILD project (Collins Birmingham University International Language Database), laying the theoretical foundation for corpus linguistics. He introduced the model of the extended unit of meaning, which posits that linguistic meaning arises from multiple layers of co-occurrence—including lexical collocation, positional relations, grammatical patterns, and contextual features—emphasizing that language functions as an organic whole.

In the field of collocational research, Sinclair significantly expanded Firth's theoretical framework in two key directions. First, he introduced the notion of "positional parameters", arguing that lexical co-occurrence should be analyzed not only in terms of frequency but also in terms of positional distribution—specifically, whether collocates appear to the left or right of the node word. This distinction underpins the functional differentiation between *positionally restricted collocations* (typically involving grammatical words such as articles and prepositions) and *positionally free collocations* (often comprising lexical items such as nouns and verbs). Second, Sinclair proposed the concept of colligation to describe the co-occurrence patterns of grammatical words, thereby establishing a clear contrast with collocation, which pertains to lexical items. This distinction addresses the theoretical gap in Firth's model regarding the blurred boundary between grammatical and lexical co-selection.

More critically, Sinclair proposed a dual framework comprising the open-choice principle and the idiom principle. The former suggests that language users, in theory, can freely combine lexical items, while the latter reveals that, in actual usage, meaning is more often conveyed through fixed collocations and patterned expressions. This framework was further enriched by the theory of upward and downward collocation. Downward collocation takes high-frequency words as node terms and examines their low-frequency collocates to uncover domain-specific semantic features. In contrast, upward collocation uses low-frequency lexical items as nodes and analyzes their co-occurrence with high-frequency grammatical words to construct stable grammatical frames.

These theoretical innovations not only deepen our understanding of language structure but also advance the development of Pattern Grammar, providing a robust methodological foundation for empirical research into linguistic regularities within specialized registers.

2. Pattern Grammar

Pattern Grammar is rooted in the British tradition of descriptive linguistics, which places strong emphasis on the study of syntactic patterns. Its historical lineage can be traced back to the 1960s, particularly to Hornby's pioneering investigation into English verb patterns and the subsequent publication of *A Guide to Patterns and Usage in English* (Hornby, 1954), which laid the groundwork for systematic pattern-based analysis in English grammar.

Sinclair's (1991) emphasis on corpus data laid the foundation for corpus-driven linguistic description. He also advanced several influential linguistic claims, including the inseparability of lexis and grammar, the close relationship between meaning and pattern, and the notion that co-selected patterns constitute the primary units of meaning. Following the compilation of the Collins COBUILD series on grammatical patterns (Francis, Hunston & Manning, 1996, 1998), Hunston and Francis (2000) systematized these findings into the framework of Pattern Grammar, offering a comprehensive account of its core concepts, theoretical foundations, and applied value.

According to Hunston and Francis (ibid.), a pattern is defined as a relatively fixed phraseological unit governed by a verb,

adjective, or noun, and composed of words that frequently co-occur with it—including prepositions, phrases, and clauses (Hunston & Francis, 2000: 3). Patterns encompass traditional grammatical relationships such as transitive verb constructions, noun and adjective complementation, and the sequencing of prepositional phrases. These patterns can be identified through corpus-driven methods, particularly via keyword-based concordance searches and distributional analysis.

In short, patterns are observable regularities in corpus-based usage, and their identification relies on criteria such as frequency, lexical co-selection, and meaning dependency.

Building on the theoretical foundations laid by Firth and Sinclair, Francis articulated two core principles that form the cornerstone of the Pattern Grammar framework. First, lexis and grammar are inseparable. In contrast to traditional grammar, which treats vocabulary as mere fillers of pre-established grammatical slots, Pattern Grammar posits that the basic unit of language is grammaticalized lexis rather than lexicalized grammar. In other words, lexical choice and grammatical patterning occur simultaneously in a dynamic process, rather than following a sequential model in which grammar precedes and governs lexical insertion. Second, patterns are intrinsically linked to meaning. Grammatical patterns are not merely surface-level structural arrangements; they function as fundamental units of semantic construction. A given pattern tends to encode a specific semantic function, while different patterns correspond to distinct semantic domains. This form–meaning association is both stable and generalizable, providing a principled basis for empirical analysis.

The encoding system of Pattern Grammar is guided by three core principles: flexibility, transparency, and consistency (Hunston & Francis, 2000: 33). It rests on two foundational premises: (1) grammar should be described based on actual language use, with concordance lines from corpora serving as the empirical basis for identifying word patterns; and (2) patterns associated with a given lexical item can be observed through extensive corpus evidence (Hunston & Francis, 2000: 250).

Compared to traditional grammatical annotation, Pattern Grammar introduces a more radical design. It employs a combination of abbreviated grammatical labels and specific lexical items or descriptive elements—for example, V n as n. Each pattern consists of three main components: the core lexical item, complementation elements, and additional constituents. The core item (e.g., V for verb) is the essential anchor of any pattern and cannot be omitted. Complementation and other elements are appended to the core item depending on usage conventions. These may take the form of specific words (e.g., V so / not, as in I hope so), structural descriptors (e.g., V n to-inf, as in They persuaded him to leave), or a combination of both (e.g., V n at n, as in He filled the bottle with water). By combining these components as needed, researchers can systematically represent the diverse syntactic patterns associated with a given lexical item.

To date, research on Pattern Grammar has predominantly focused on English, both in domestic and international scholarship, while investigations of pattern grammar structures in French remain relatively scarce. This study contributes to advancing the exploration of Pattern Grammar in French by offering empirical insights into its structural and semantic manifestations within a specialized domain.

3. Research design

3.1 Research problems

This study focuses on the high-frequency "N à N" structure in French, aiming to identify its grammatical patterns, classify its semantic functions, and analyze its distribution across a

specialized corpus. Two research questions are addressed:

(1) What are the recurring "N à N" patterns in technical French, and how frequently do they occur ?

(2) What types of semantic relations are encoded by the "N à N" structure in specialized discourse, and do these relations form stable semantic categories?

3.2 Introduction of the Corpus

This study draws on a self-compiled corpus of electricity-related French developed by Yang Tong (2024). The corpus comprises texts from a range of professional sources, including contracts and agreements, construction manuals, conference papers, and academic exchanges in the electricity sector. It covers twelve representative subfields within electrical engineering, as shown in Table 1. According to corpus statistics, the dataset contains 462,737 tokens and 36,269 types.

3.3 Methodology

A defining feature of Pattern Grammar is its reliance on a corpus-driven approach, which differs fundamentally from the corpus-based approach (Tognini-Bonelli, 2001). In corpus-driven research, grammatical and semantic categories are not pre-imposed but are instead inductively derived from recurrent linguistic patterns observed in authentic language use.

This study employs AntConc 4.3.1 to analyze a specialized corpus, with a focus on identifying and examining actual instances of the "N à N" structure and their semantic characteristics. The procedure consists of several steps. First, candidate patterns were preliminarily extracted using the keyword query "à *", followed by manual verification to ensure that each instance conforms syntactically to the "N à N" structure and exhibits semantic relevance. Second, AntConc's keyword-in-context (KWIC) and collocation functions were used to calculate the frequency of specific patterns, analyze their contextual co-occurrence features, and observe semantic tendencies within the electricity-related corpus. Subsequently, semantic categorization was conducted based on the lexical types of the two nouns involved in each "N à N" structure and their combinatory configurations. The frequency of each semantic pattern was also recorded. For cases with ambiguous semantic boundaries, contextual interpretation was applied to enhance the accuracy and consistency of classification.

Table 1: sources of self-compiled corpus

twelve representative subfields	Digital Electronics (électronique numérique)
	Energy Conversion Systems (systèmes de conversion d'énergie)
	Signal Processing (traitement du signal)
	Analog Circuits (électronique analogique)
	Photovoltaic Energy (énergie photovoltaïque)
	Simulation Tools and Thermal Applications in Conversion (outils de simulations et applications thermiques en conversi)
	Modeling and Dimensioning of Synchronous Actuators (modélisation et dimensionnement d'un actionneur synchrone)
	High-Voltage Dielectric Materials and Components (matériaux et composants diélectriques – haute tension)

	Operational Safety (sureté de fonctionnement)
	Energy Conversion Systems for Embedded Applications (système de conversions d'énergie pour applications embarquées)
	Reliability of Components and Systems (fiabilités des composants et systèmes)
	Computer Engineering for the EEA Sector (génie informatique pour l'EEA)

4. Research results and analyse

4.1 Frequency Analysis

Given the large volume of data, the study selected the top 30 structurally valid "N à N" patterns based on extraction results. Each pattern was annotated for its semantic function and ranked in descending order of frequency, as shown in Table 2. The most frequent instances of the "N à N" structure in the electricity corpus include *pompe à chaleur* (228 occurrences), *pile à combustible* (184), *gaz à effet de serre* (75), *piles à combustible* (61), and *stockage à air comprimé* (55). These results indicate that the "N à N" structure is relatively common in electricity-related French.

Table 2: Frequency of "N à N" patterns

No.	Examples	Frequncy
1	pompe à chaleur	228
2	pile à combustible	184
3	gaz à effet de serre	75
4	stockage à air comprimé	61
5	machine à réluctance variable	55
6	déchiqueteuse à disque	45
7	électrolyseur à puissance nominale	15
8	prédiction à court terme	15
9	machine à courant continu	13
10	tension à vide	13
11	combustible à hydrogène	13
12	onduleur à quatre bras	12
13	Turbine à gaz	11
14	bus à courant continu	10
15	machine synchrone à aimants permanents	8
16	fonctionnement à charge partielle	6
17	déchiqueteuse à rotor	6
18	système à air comprimé	5
19	gaz à cycle combiné	5
20	machine à vapeur	5
21	moteur à combustion interne	5
22	machine à réluctance variable à double saillance (MRVDS)	5
23	moteur à aspiration naturelle	4
24	charge à l'anode	4
25	signal à bruit	4
26	éolienne à vitesse variable	4

27	éolienne à axe horizontal	3
28	fonctionnement à vitesse variable	3
29	décharge à courant constant	3
30	chaleur à basse température	3

The extracted instances reveal the presence of three structural types: standard patterns, nested patterns, and extended patterns. The standard pattern corresponds to the basic "N à N" structure. Nested patterns exhibit more complex configurations such as "N à N ADJ à ADJ N", while extended patterns include various variants such as "N à N ADJ", "N à ADJ/NUM N", and "N ADJ à N ADJ". As shown in Table 2, the corpus contains 11 standard patterns (36.7%), 18 extended patterns (60%), and 1 nested pattern (3.3%). These figures suggest a tendency toward structural complexity in technical discourse.

The electricity corpus represents a highly specialized register, in which terminology is required to precisely convey structural, parametric, and functional information. To meet these communicative demands, patterns frequently incorporate modifiers such as adjectives, compound nouns, and prepositional phrases, resulting in nested patterns. Pattern Grammar exhibits high extensibility in specialized domains, enabling it to accommodate increasingly complex semantic structures.

4.2 Semantic categorization of patterns

Shared semantics refers to the phenomenon in which multiple terms exhibit semantic commonality and can be grouped into the same category based on similar functional, structural, or conceptual properties. The co-occurrence of lexical items with their grammatical patterns is not arbitrary; rather, it is systematically linked to specific meanings.

Based on the semantic relationships between the constituents of each pattern, "N à N" patterns can be considered "N1 à N2". With "N1" as the semantic head, can be broadly classified into three categories: technical equipment, energy and storage, and parameter/control, as shown in Table 3.

Table 3: Semantic category of examples

Semantic category	Examples
Technical equipment	pompe à chaleur machine à réluctance variable machine à réluctance variable à double saillance machine à courant continu machine synchrone à aimants permanents machine à vapeur moteur à combustion interne moteur à aspiration naturelle déchiqueteuse à disque électrolyseur à puissance nominale onduleur à quatre bras turbine à gaz éolienne à vitesse variable éolienne à axe horizontal déchiqueteuse à rotor

Semantic category	Examples
Energy and storage	combustible à hydrogène pile à combustible gaz à effet de serre gaz à cycle combiné stockage à air comprimé système à air comprimé
Parameter/control	charge à l'anode tension à vide fonctionnement à charge partielle fonctionnement à vitesse variable décharge à courant constant signal à bruit bus à courant continu chaleur à basse température prédiction à court terme

Further semantic distinctions can be made within each major category based on the lexical and conceptual roles of "N1" and "N2", as illustrated in Table 4. For example, in the technical equipment category, terms such as *machine à vapeur*, *moteur à combustion interne*, and *turbine à gaz*, all feature "N2" as an energy source that provides power to the "N1" device. These constructions share a common underlying cognitive pattern: the postposed "à N2" phrase serves to specify and assign a technical parameter to the core noun "N1". However, while this parameterization relation is structurally unified, its semantic realization varies depending on the category of "N1".

Table 4: the conceptual roles of "N1" and "N2"

Semantic category	Subcategory	Examples
Technical equipment	N2 functions as the energy source or driving force for N1	machine à réluctance variable
		machine à réluctance variable à double saillance
		machine à courant continu
		machine synchrone à aimants permanents
		machine à vapeur
		moteur à combustion interne
		machine à réluctance variable
		turbine à gaz
		pompe à chaleur
	N2 specifies the technological mechanism or design principle of N1	moteur à aspiration naturelle
		onduleur à quatre bras
		éolienne à vitesse variable
		éolienne à axe horizontal
		déchiqueteuse à rotor

		déchiqueteuse à disque
	N2 describes the state of N1	électrolyseur à puissance nominale
Energy and storage	N2 as the energy of N1	combustible à hydrogène
		pile à combustible
	N2 as the medium of N1	stockage à air comprimé
		système à air comprimé
		gaz à effet de serre
	N2 describes the state of N1	gaz à cycle combiné
Parameter/control	N2 describes the state of N1	charge à l'anode
		tension à vide
		fonctionnement à charge partielle
		fonctionnement à vitesse variable
		décharge à courant constant
		bus à courant continu
		chaleur à basse température
		prédiction à court terme
	N1 and N2 define a contrastive relation	signal à bruit

It is worth noting that the productivity of the "N à N" structure in technical terminology is not limited to expressing parameterization. A key example is the term *signal à bruit*, which reveals another important semantic category: relational opposition and integration. Unlike phrases such as *pompe à chaleur* or *tension à vide* , where the "à N2" component modifies the core noun "N1" by specifying a technical parameter, *signal à bruit* does not reflect a modifier-head relationship. Instead, signal and bruit function as a pair of interdependent, contrastive elements. The meaning of the phrase is not "a signal characterized by noise", but rather "the ratio between signal and noise"—a new abstract concept defined by the quantitative and functional relationship between the two nouns. This demonstrates that the "N à N" construction is not restricted to attributive or descriptive functions; it can also serve to construct relational entities, where meaning emerges from the interaction between components rather than from hierarchical modification.

5. Conclusion

Through the pattern identification and semantic categorization of "N à N" constructions in the electricity-related corpus, this study reveals the complexity and systematicity of pattern expressions in scientific and technical French. Pattern Grammar not only captures the structural regularities of high-frequency terminology, but also identifies nested pattern and semantic variants, highlighting the close relationship between linguistic form and specialized meaning.

The findings indicate a tendency toward increased precision and semantic clarity in pattern expressions within technical registers. Pattern Grammar demonstrates broad potential for applications in terminology extraction, semantic modeling, and register-based analysis. Future research may extend this approach to other registers and patterns, contributing to the development of a more comprehensive constructional framework for scientific language.

References

[1]. Hornby, A. (1954). *A guide to patterns and usage in English*. Oxford University Press.

[2]. Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins.

[3]. Hunston, S. (2002/2006). *Corpora in applied linguistics*. WPC.

[4]. Firth, J. R. (1957). Modes of meaning. In J. R. Firth (Ed.), *Papers in linguistics* (pp. xx–xx). Oxford University Press.

[5]. Francis, G., Hunston, S., & Manning, E. (1996). *Collins COBUILD grammar patterns 1: Verbs*. Harper Collins.

[6]. Francis, G., Hunston, S., & Manning, E. (1998). *Collins COBUILD grammar patterns 2: Nouns and adjectives*. Harper Collins.

[7]. Sinclair, J. M. (1987). Grammar in the dictionary. In J. M. Sinclair (Ed.), *Looking up: An account of the COBUILD project in lexical computing* (pp. xx–xx). Collins ELT.

[8]. Sinclair, J. M., & Renouf, A. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. xx–xx). Longman.

[9]. Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford University Press.

[10]. Sinclair, J. M. (2000). Lexical grammar. *Naujoji Metodologija*, *xx*, xx–xx.

[11]. 陈功, & 梁茂成. (2017). 型式语法的产生、特点及其应用价值 [The emergence, characteristics, and application value of pattern grammar]. *外语学刊*, (01), 17–24. https://doi.org/10.16263/j.cnki.23-1071/h.2017.01.004

[12]. 苏杭, & 卫乃兴. (2023). 型式语法、构式语法、局部语法：汇融与歧异 [Pattern grammar, construction grammar, and local grammar: Convergence and divergence]. *外语教学与研究*, *55*(06), 925–937, 961. https://doi.org/10.19923/j.cnki.fltr.2023.06.014

[13]. 王勇. (2008). 行走在语法和词汇之间——型式语法述评 [Walking between grammar and lexis: A review of pattern grammar]. *当代语言学*, (03), 257–266, 286.

[14]. 杨桐. (2024). 基于语料库的电力法语词汇教学研究 [Corpus-based study on French vocabulary teaching in the electricity field]. *华北电力大学学报(社会科学版)*, (02), 108–116. https://doi.org/10.14092/j.cnki.cn11-3956/c.2024.03.029