論文

A Contrastive Analysis and Rrelated Problems of Information Contents in Japanese and French Abstracts of Chemistry

Sophie PALVADEAU
University of the Ryukyus

This paper aims to compare information contents in Japanese and French abstracts of chemistry written by native speakers. A comparison of the informational content and its degree of completeness is drawn for each different abstracts from *one-segment* to *four-segments*. The very first point revealed by this segmentation was the trend in French corpora to summarize in two or three "*informative modules*" whereas in the Japanese corpus, this is done in three or four modules. Then, *single-segment abstracts* are only attested in our french corpora and may be compared to a short advertising message. Number of segments is a prevalent parameter shaping the informative sequences, consequently it causes problems to find comparable features in a cross-linguistic perspective.

1. Introduction

This communication aims to compare the contents of Japanese and French abstracts of chemistry written by native speakers. The choice of "authors' abstracts" has been motivated by a previous study's results[1] that had shown the major role of "guidelines for contributors" in the shaping of articles. Because required characteristics for "full articles", "notes" or "communications" basically depend on each Journal's policy, it appeared difficult to compare these articles. Then, we formulated the preliminary hypothesis that corpora of abstracts may share at least one property - to provide a « short statement that gives the salient features and drawing attention to the novel aspect of a research paper ». However, it was clear that this *identity of purpose* didn't match with « a single written pattern »: according to instructions to authors in the Japanese corpus[2], specific content[3] was requested for the abstracts whereas French corpora[4] only focused on number of words. Although it was noted, Japanese and French scientific writing guidelines[5], that two main patterns coexisted from the 1950s, it was added that these were not « strictly observed ». These guidelines explained the use of IMRAD's plan to shape an abstract perceived as a « mini-version » of a full paper, and mentioned the « structured abstract » defined as « readers-focused » text with every thematic parts clearly labelled

[1] cf.references: Palvadeau 2001.

[2] « *Nippon Kagaku kaishi* » 1998 [NKK] (104 abstracts).

[3] « The Japanese abstracts must included the essential of the objectives, method and results in order to provide a clear idea of the full paper content » 「和文要旨はその論文の主旨が把握できるように、目的、方法、結果を要約し記述する。」

[4] « *Comptes rendus de l'Académie des sciences-chimie* » 1998 [CRC] (88 abstracts) and the on-line « *Lettre du Département des Sciences Chimiques du CNRS* » 1997-2000 [LSC] (48 abstracts).

[5] cf.references: Desjeux 1997; Kinoshita 1994; Tanaka 1998; Yamazaki 1998.

(« objective », « setting », « results », etc).

Besides the "global tagging" made possible by "Introduction-Method-Result And Discussion" (IMRAD) labels, we defined our objectives in order to describe the content of the information expressed in Japanese and French abstracts of chemistry, to evaluate their completeness and the modality adopted by authors in relation to various "external" parameters ("category of the attached article", "type of authors' laboratory"). After a presentation of the methodology used and problems encountered, main results will be discussed.

2. Methodology

The first step has consisted in defining the *criteria of segmentation* adapted to our corpora. Focused on *informational content* combined in the field of chemistry, what we called "*informative module*" has been consequently defined as a *container* of "one message in one or more sentences" or like "objectives and related operations" corresponded to "a single sentence conveying more than one message". At least nine *informative modules* had been identified, as follows.

Figure 1: « Informative modules from the Japanese and French authors abstracts of chemistry ».

IMRAD pattern	Code	Categories of information
Introduction	M1	Reminder of information, knowledge (field background) as the starting point for the authors' research presented in the paper.
	M2	Expression of motivations related with a scientific or economic problematic.
	M3	Objectives or paper contents.
Method	M4	Method (principles adopted) or procedure description
Results	M5	Method (or process) with related results
	M6	Results
Discussion	M7	Discussion
	M8	Conclusion or claim on results
	M9	Miscellaneous : equipment, sponsorship, etc.

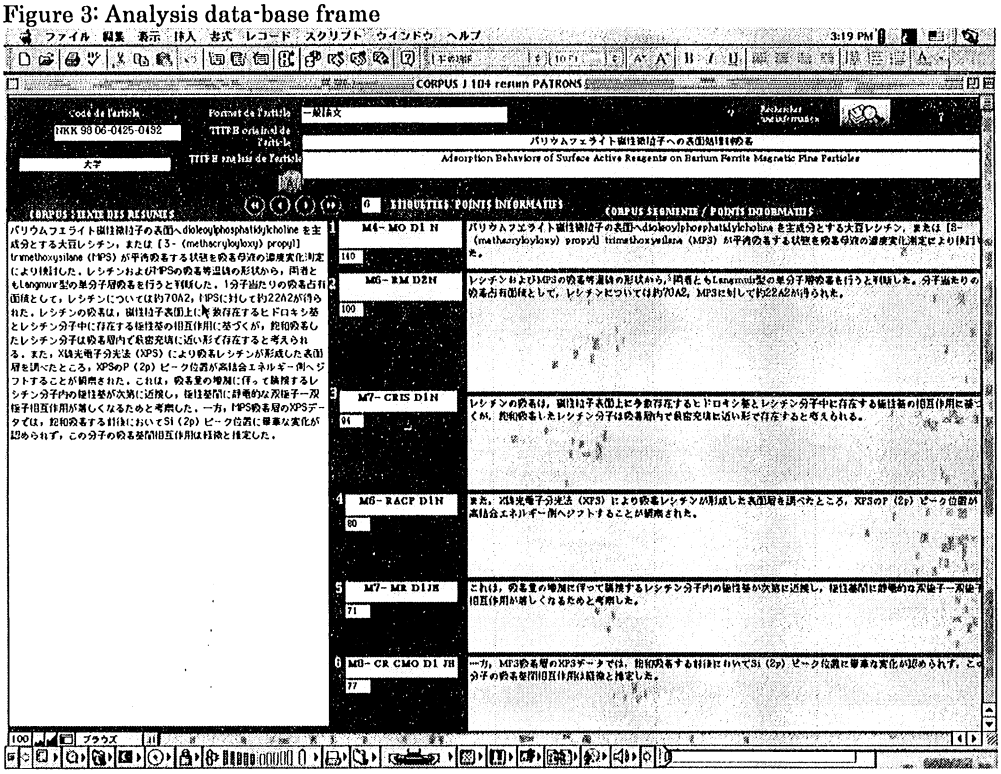
In order to characterize the completeness of the above *informative modules* we had to create a scale based on criteria easy to identify whatever the topic. From a « general description » to a more and more « complete one », the next three levels (Fig.2) were determined.

Figure 2: « The Three levels of the scale of completeness »

General description [DG]	Information of degree 1 [D1]	Information of degree 2 [D2]
Giving the name of the reaction mechanisms involved (ex: addition, condensation, substitution, etc)	[DG] + name of its reactants	[D1] + operating conditions data (temperature, etc)
Naming the properties of a component	[DG] + data regarding its molecular structure (shape)	[D1] + related measured data (interatomic distances, etc)
General principles, mechanism of a setup equipment of chemical engineering	[DG] + specific applications for a particular product	[D1] + related numerical data

Encoding of modality was realized as a binary system: « neutral mode »[N] vs. « modal clause » [JE]. The first code marked sentences written in neutral forms and the second one involved sentences containing evaluation or judgment by the author(s). We also registered the type of articles attached to the abstracts. Our mini-database has been realized on *File Maker Pro 5 ®* as follows (figure 3) :

Figure 3: Analysis data-base frame



At the end of this step, we had to deal with more than *fifty labels* provided by our analysis. This meant we had difficulties in the presentation of the results (linear list of categories) and a challenge to keep clearly visible the "shape" of the entities named abstracts in a cross-linguistic comparison perspective. The pragmatic approach adopted was to study separately "single segment" then "two-segment" to "six-segment" abstracts of both Japanese and French corpora. In addition, we tried to describe the relation between location of "informative modules" and « completeness » or « modality » variations in relation to the nature of the attached full paper.

3. Results-Discussion

The very first point revealed by this segmentation was the trend in French corpora to summarize in two or three *"informative modules"* whereas in the Japanese corpus, this is done in three or four modules, as seen in figure 4.

Figure 4 : Number of *segments* in the Japanese and French corpora.

Corpus (nb of abstract)	Number of "informative modules" segments (% of abstract)					
	1	2	3	4	5	6
CRC (88)	6 (7%)	36 (41%)	33 (38%)	7 (8%)	5 (6%)	0
LSC (48)	2 (4%)	15 (31%)	15 (31%)	11 (23%)	3 (6%)	2 (4%)
CRC + LSC (136)	8 (6%)	51 (38%)	48 (35%)	18 (13%)	8 (6%)	2 (1,5%)
NKK (104)	0	6 (6%)	42 (40%)	39 (38%)	14 (14%)	3 (3%)

3.1. « Single segment » abstracts

This type of abstract was only attested in our French corpora (about 6%) with six samples from CRC and two samples from the *on-line* LSC as follows (figure 5).

Figure 5 : *Informative modules* of "one-segment abstracts".

IMRAD	Informative modules	corpus code	attached article type	nb.of cases
Introduction	M2-S D1 JE : "Scientific motivations-problematics" [Completeness degree 1] ["modality included"]	CRC	notes	1
	M3-OBgDG OP N: "Global objectives and related operation" ["general description"] – [neutral]	CRC	short paper	1
Method	M4-MA D1N : "Applied method- Approach" [completeness degree1] – [neutral]	CRC	notes	1
			short paper	1
Results	M5-MA DGN CP DGN: "Applied method-Approach" ["general description"] – [neutral] + "result characteristics" ["general description"] – [neutral]	LSC	*nouvelle scientifique*	1
	M5- MO D2N RI D1N : "Procedure" [completeness degree 2] + "end-product identity" [Completeness degree 1] – [neutral]	CRC	notes	1
	M6- RLIPA-D1N: "Results linking experimental approach with end-product identity" [completeness degree1] – [neutral]	CRC	short paper	1
	M6- RC DG JE: "End-products characteristics" ["general description"] ["modality included"]	LSC	*nouvelle scientifique*	1
	M8- CR CMO JE: "Conclusion claiming knowlegde regarding studied mechanism" ["general description"] ["modality included"]	CRC	notes	1

"Results" and "conclusion" generated the main content of the "one segment abstracts." Neutrality of the sentences is observed in five cases out of eight. In addition, "general description" appeared to be prominent in four samples, then three cases expressed "degree one of completeness" for only one "degree two". The above *M2- M3-* and *M4-* samples may be compared to a short advertising message telling people to read the *attached article* rather than to provide a digest of the full text.

3.2. « Two segment » abstracts

As it can be noticed the above (fig.4), this category is the more important in the French corpora (38%) but only a few cases (6%) were registered in the Japanese corpus. This meant we had to compare *fifty-one* French abstracts with only *six* Japanese abstracts. We computed each pair of segments [Mn-Mn*], as follows:

Figure 6: Mapping of "two-segment" *informative modules*

Corpora	[M1-M*]	[M2-M*]	[M3-M*]	[M4-M*]	[M5-M*]	[M6-M*]	[M8-M*]
CRC (36)	8 (22 %)	1 (3%)	0	11 (30%)	7 (19%)	6 (16%)	4 (11%)
LSC (15)	6 (40%)	2 (18%)	1 (7%)	1 (7%)	0	3 (20%)	2 (13%)
NKK (6)	1 (16%)	1 (16%)	1 (16%)	50% (3)	1 (16%)	0	0

CRC and NKK's two segments abstracts mainly began with the "*M4-module*" (which expressed "method or procedure's description"), while in the *on-line* journal LSC,it was the "*M1-module*" containing "reminder of information, field background" that was prevalent.

Variation of modality was registered as follows:

Figure 7 : Modality variation in two-segment informatives modules

Variation	Corpus	Σ	[M1-*]	[M2-*]	[M3-*]	[M4-*]	[M5-*]	[M6-*]	[M8-*]
N-N	CRC	21	3	0	0	7	5	5	1
	LSC	5	0	2	1	1	0	0	1
	NKK	2	0	0	0	1	1	0	0
N-JE	CRC	7	1	0	0	3	0	1	2
	LSC	4	1	0	0	0	0	2	1
	NKK	3	0	0	1	2	0	0	0
JE-N	CRC	7	2	1	0	1	2	0	1
	LSC	6	5	0	0	0	0	1	0
	NKK	1	1	0	0	0	0	0	0
JE – JE	CRC	2	2	0	0	0	0	0	0

[N: neutral clause – JE : modal clause]

In CRC, the above results shown, perhaps consequently with the high score of seven *M4-modules* (Method), and five M5- or M6- modules, that the constant neutrality of expression [N-N] was a prevalent pattern [21cases]. Then we noted that [N-JE] and [JE-N] patterns obtained the equivalent scores of 7 cases each and surprisingly that two samples contained the [JE-JE] pattern. In the on-line journal LSC, [JE-N] was the main pattern provided six cases but didn't contrasted clearly with the others scores of five or four cases respectively with [N-N] and [N-JE] patterns.

In NKK, the small dimension of our sample made difficult the description of a trend figured by the computed three samples of [N-JE] pattern out of the six registered.

3.3. « Three segment » abstracts

Both Japanese and French corpora contained "three-segment" of *informative module*, which represented 42 abstracts of NKK (40%), 33 of CRC (38%) and 15 of LSC (31%) as seen in figure 4.

Only five sequences were similar in Japanese and French corpora, these started with M4- *informative modules.* Then three more examples of parallel pattern were noticed in M1-, M5- and M6- modules as follows.

Figure 8 : Parallel sequences of "three-segment" Japanese and French abstracts

Sequences	CRC	LSC	NKK
[M1-M5-M8]: Reminder of Information + method-results + conclusion-claim	×	○	○
[M4-M4-M8]: Method-procedure + method-procedure + conclusion-claim	○	×	○
[M4-M5-M6]: Method-procedure + method-results + results	×	○	○
[M4-M6-M4]: Method-procedure + results + Method-procedure	○	×	○
[M4-M6-M6]: Method-procedure + results + results	○	×	○
[M4-M6-M7]: Method-procedure + results + discussion	○	×	○
[M4-M6-M8]: Method-procedure + results + conclusion-claim	○	×	○
[M4-M7-M8]: Method-procedure + discussion + conclusion-claim	○	×	○
[M5-M6-M6]: Method-results + results + results	○	×	○
[M6-M5-M8]: Results + method-results + conclusion-claim	×	○	○

In other words, only seven three-segment abstracts of NKK and CRC had the same sequences and three samples of NKK and LSC had the same *informative modules*. Curiously, no common pattern was attested between the two French corpora. These features tend to strength the idea of the predominance of the characteristics of the full-articles in the shaping of the abstracts and motivate a description of the most frequently used modules for each review (figure 9, above).

Figure 9 : Scores of the most frequently used modules in three segment abstracts with scale of completeness.

Sequence number:	one			two			three		
NKK (42 abstracts)	M4 (26) 62%	DG	3 (12%)	M6 (20) 48%	DG	4 (20%)	M8 (22) 52%	DG	6 (27%)
		D1	19 (73%)		D1	4 (20%)		D1	12 (55%)
		D2	4 (15%)		D2	12 (60%)		D2	4 (18%
CRC (33 abstracts)	M4 (18) 55%	DG	2 (11%)	M6 (18) 55%	DG	2 (11%)	M8 (16) 48%	DG	7 (44%)
		D1	14 (78%)		D1	10 (56%)		D1	8 (50%)
		D2	2 (11%)		D2	6 (33%)		D2	1 (6%)
LSC (15 abstracts)	M1 (7) 47%	DG	4 (57%)	No predominant score			M8 (7) 47%	DG	4 (57%)
		D1	3 (43%)					D1	3 (43%)

Both NKK and CRC three segment abstracts have globally the same categories of information: M4 (*method or procedure description*) , M6 (*results*) and M8 (*conclusion or claim on results*). Regarding the *scale of completeness*, with parallel trends in Japanese and French corpora, the D1 type appears to be the most current degree for M4 (73% NKK; 78% CRC) and M8 (55%NKK, 50%CRC). On the contrary divergence is noticed here with 60% of the results (M6) expressed in Japanese content numerical data (D2 degree) while in their French equivalent the score is only 33% of D2 (56% of D1).

Then, the on-line LSC abstracts differ

from both NKK and CRC in two ways : its first sequence is a *reminder of information* and no numerical data are expressed.

3.4. « Four segment » abstracts

Meanwhile "four segment" abstracts shape 38% of NKK abstracts, the French corpora provides respectively 8% (CRC) and 23% (LSC) of this category. Then because no parallel pattern of informative modules has been observed, only a concise cross-language comparison focused on the *scale of completeness* is presented figure 10.

Figure 10 : Scale of completeness in "four segment" abstracts modules.

Sequence number	degree	NKK (39 abstracts)	CRC (7 abstracts)	LSC (11 abstracts)
One	M*DG	18% (7)	43% (3)	36% (4)
	M*D1	66% (26)	43% (3)	36% (4)
	M*D2	15% (6)	14% (1)	27% (3)
Two	M*DG	10% (4)	29% (2)	36% (4)
	M*D1	56% (22)	29% (2)	45% (5)
	M*D2	33% (13)	42% (3)	18% (2)
Three	M*DG	2.5% (1)	43% (3)	18% (2)
	M*D1	54% (21)	29% (2)	63% (7)
	M*D2	44% (17)	29% (2)	18% (2)
Four	M*DG	8% (3)	43% (3)	45% (5)
	M*D1	72% (28)	43% (3)	45% (5)
	M*D2	20% (8)	14% (1)	9% (1)

As these scores do not include the type of information provided by modules, it is clear that only a basic picture can be drawn here. Degree one (D1) is the most common level of information written in NKK 's four segment abstracts whatever the sequence number. Meanwhile *sequence three* have a 44% score of D2 (numerical data included) and few DG are noticed. French CRC corpora' trend seems to depend more on the *sequence number*: DG or D1 (one) then D2 (two), DG (three) and DG or D1. LSC's abstracts generally have a D1 level whatever the module's location.

In order to describe more precisely the NKK four segments abstracts we have computed their most used informative modules as follows.

Figure 11 : Most frequent informative modules of four-segments japanese abstracts.

sequence number	one	two	three	four
module code	M4	M6	M7	M8
score	24 (61.5%)	15 (38%)	14 (36%)	28 (72%)

As it was observed with the three-segments abstracts, M4 (method-procedure), M6 (results) and M8 (conclusion-claim) are the most used categories of information. M7 (discussion) is also commonly located at the sequence number three. If we compare the 38% of M6 with the score of 48% recorded from three-segment modules (figure 9) we may explained the gap as a result of a diversification of informative modules linked with the number of segment expressed.

4. Conclusion

In this paper we compared the informative modules that described the content of

author's chemical abstracts. The trend to figure this content with "two or three segments" was observed in French whereas the Japanese corpus shown the use of "three or four segments". Besides this appearance of similarity, it turned out that only few sequences were attested both in the two languages, at least we reported the number of ten three-segment informative modules. Even in such instance, different types of informational content was involved. At least, by using the same tools mapping "what it is conveyed" in sentences of chemistry, we started to point out the diversity of expressions involved in our corpora.

Acknowledgements:

We are grateful to Professor Vivian NOBES (University of Tsukuba, College of Nursing and Medical Technology) for her assistance in checking the English grammar and spelling of this paper. (2001,December)

References:

[1] Desjeux M-F., Mary J-Y. and Desjeux J-F 1997."Guide pratique de la communication scientifique". Ellipses. Paris.

[2] Georgakopoulou A and Goutsos D. 1999."Discourse Analysis- An introduction". Edinburgh University Press.

[3] Halliday M.A.K and Martin J.R. 1993. "Writing Science".University of Pittsburgh Press.

[4] Palvadeau S. 2001. "Le format des articles scientifiques, comparaison et traduction des usages observés sur corpus japonais, anglais et français". 琉球大学欧米文化論集第、45 号 2001 年 3 月 (51-71) .

[5] Robert D. A. and Bouillaguet A. 1997. "L'analyse de contenu";Que-sais-je? N°3271. Presse Universitaire de France.

[6] 木下雄 1994 年「レポートの組み立て方」第１５版、筑摩書房

[7] 田中潔 1998 年「手ぎわよい科学論文の仕上げ方」第２版、共立出版 KK.

[8] 山崎茂明 1997 年「生命科学論文投稿ガイド」第２版、中外医学社

著者紹介

Sophie PALVADEAU (正会員)

フランス国立東洋言語文化学院 (INALCO) 日本語学博士課程研究免状(DEA).

現在、琉球大学法文学部国際言語文化学科フランス語学助教授.

翻訳の面から見る日本語・英語・仏語の理系の文章・ターミノロジーに興味を持つ.

フランス日本研究会、日本の化学史学会等会員.