Journal of Digital Imaging (2022) 35:993–1007
https://doi.org/10.1007/s10278-022-00619-6

ORIGINAL PAPER

What Influences the Way Radiologists Express Themselves in Their Reports? A Quantitative Assessment Using Natural Language Processing

Amandine Crombé[1,2] · Mylène Seux[1] · Flavie Bratan[1,3] · Jean-François Bergerot[1,4] · Nathan Banaste[1,5] · Vivien Thomson[1,6] · Jean-Christophe Lecomte[1,7,8] · Guillaume Gorincour[1,9]

Received: 14 October 2021 / Revised: 7 March 2022 / Accepted: 9 March 2022 / Published online: 22 March 2022

Abstract

Although using standardized reports is encouraged, most emergency radiological reports in France remain in free-text format that can be mined with natural language processing for epidemiological purposes, activity monitoring or data collection. These reports are obtained under various on-call conditions by radiologists with various backgrounds. Our aim was to investigate what influences the radiologists' written expressions. To do so, this retrospective multicentric study included 30,227 emergency radiological reports of computed tomography scans and magnetic resonance imaging involving exactly one body region, only with pathological findings, interpreted from 2019–09-01 to 2020–02-28 by 165 radiologists. After text pre-processing, one-word tokenization and use of dictionaries for stop words, polarity, sentiment and uncertainty, 11 variables depicting the structure and content of words and sentences in the reports were extracted and summarized to 3 principal components capturing 93.7% of the dataset variance. In multivariate analysis, the 1st principal component summarized the length and lexical diversity of the reports and was significantly influenced by the weekday, time slot, workload, number of examinations previously interpreted by the radiologist during the on-call period, type of examination, emergency level and radiologists' gender (*P* value range: < 0.0001–0.0029). The 2nd principal component summarized negative formulations, polarity and sentence length and was correlated with the number of examination previously interpreted by the radiologist, type of examination, emergency level, imaging modality and radiologists' experience (*P* value range: < 0.0001–0.0032). The last principal component summarized questioning, uncertainty and polarity and was correlated with the type of examination and emergency level (all *P* values < 0.0001). Thus, the length, structure and content of emergency radiological reports were significantly influenced by organizational, radiologist- and examination-related characteristics, highlighting the subjectivity and variability in the way radiologists express themselves during their clinical activity. These findings advocate for more homogeneous practices in radiological reporting and stress the need to consider these influential features when developing models based on natural language processing.

Keywords Natural language processing · Text mining · Principal component analysis · Structured reports · Emergency radiology · Teleradiology

✉ Amandine Crombé
a.crombe@imadis.fr

[1] IMADIS, 48 rue quivogne, 63002 Lyon, France

[2] University of Bordeaux, 33000 Bordeaux, France

[3] Department of Diagnostic and Interventional Imaging, Centre Hospitalier Saint-Joseph Saint-Luc, 69007 Lyon, France

[4] Ramsay Générale de Santé, Clinique Convert, 01000 Bourg-en-Bresse, France

[5] Department of Radiology, Hôpital Nord-Ouest, 69400 Villefranche-sur-Saône, France

[6] Ramsay Générale de Santé, Clinique de la Sauvegarde, 69009 Lyon, France

[7] Centre Hospitalier de Saintonge, 17100 Saintes, France

[8] Centre Aquitain d'Imagerie, 33600 Pessac, France

[9] ELSAN, Clinique Bouchard, 13006 Marseille, France

Abbreviations

COVID-19	Coronavirus disease 2019
ERR	Emergency radiological reports
mANOVA	Multiple way analysis of variance
NEPIR	Number of examinations previously interpreted by radiologist
NLP	Natural language processing
PACS	Picture archiving communication system
PC	Principal component
PCA	Principal component analysis
SW	Stop words

Introduction

Most radiological reports are organized in various sections comprising free-text components with unstructured paragraphs. These reports contain original information that can be extracted through natural language processing (NLP) and enable the automated labeling of large databases with excellent accuracy [1–3]. Thus, several potential applications have already emerged that could be of great interest for radiology departments and global organization, for instance: rapid recruitment of patients for retrospective studies; development of NLP tools to stress missed items in the conclusion compared with the results section; monitoring of radiological activity for epidemiological purposes, globally and for specific pathologies; anticipation of human and technical resources; benchmarking of the activity of a radiology department compared with others; and quality assurance, accreditation and quality assessment of radiologic practice [4–6].

However, developing NLP models based on radiological reports requires either consistency regarding the radiologists' means of expression and consideration for all possible radiological formulations. Furthermore, better understanding what influences these radiologists' means of expression could be useful to realize if societal bias and prejudices also exist in our daily practice (for instance, that 'women performs more detailed and literary reports than men', or that 'younger and less experienced radiologists perform reports with more expressions related to doubt and uncertainty and less meaningful words').

In France, emergency radiology is continuously taught during the entire radiological curriculum as a transversal core competency. Moreover, many radiologists continue to perform emergency radiology during on-call periods even after the end of their residency and fellowship. Hence, emergency radiological reports represent a rare opportunity to compare the way radiologists of various backgrounds (in terms of gender, education, experience and radiological skills) express themselves under various conditions (in terms of workload, stress, tiredness, hour or day of the week), in contrast to other specialties such as abdominal imaging, musculoskeletal imaging or women imaging, for instance.

In parallel, because of the increasing use of diagnostic imaging in emergency departments (for instance, + 42% between 2002 and 2015 in France), emergency radiology has been urged to reorganize to cope with this escalation and to be able to transmit results in real time to emergency physicians within the shortest time, 24 h a day, 7 days a week [7, 8]. Various systems have been developed, such as resident coverage with subspecialist overreading, 24-h attending coverage with initial resident interpretation or limited/no resident involvement and fully dedicated emergency teleradiology [9, 10]. In particular, the common information and technology tools used by teleradiology have made possible large multicentric radiological data collection [11].

Therefore, the aim of this study was to investigate how the context of the on-call periods and the characteristics of the radiologists could modify the structure and content of emergency radiological reports as quantified by using NLP tools, in a large multicentric emergency teleradiology cohort.

Materials and Methods

Study Design and Data Origin

This retrospective study was approved by the French Ethics Committee for the Research in Medical Imaging (CERIM) review board (CRM-2106–168) and performed in accordance with good clinical practices and applicable laws and regulations. The need for written informed consent was waived because of its retrospective nature.

We collected all the radiological reports from IMADIS Emergency Teleradiology from 2019–09-01 to 2020–02-28. This date range corresponded to a period of 6 months before the beginning of the coronavirus disease (COVID) 2019 pandemic in France (which strongly modified the activity of emergency departments), during which the experience and skills of radiologists were considered constant.

IMADIS Organization

IMADIS is a French medical company dedicated to the remote interpretation of emergency imaging examinations (magnetic resonance imaging (MRI), computed tomography (CT) scan and radiographs) for public and private hospitals ($n = 61$ during the study period). The IMADIS activity mostly occurs during night-time on-call periods (from 6 p.m. to 8 a.m.), except during the weekend where the on-call periods last 24 h.

The radiologists working at IMADIS all have another main radiological activity, either as residents or radiologists in private or public, specialized or non-specialized radiology departments.

During their activity for IMADIS, the radiologists worked in three teleradiological interpretation centers (Lyon, Bordeaux and Marseille, France) within teams of 2 to 6 radiologists per center on secured dedicated workstations. For each on-call period, the total number of radiologists was anticipated and adapted to the expected activity, with a balance between senior and junior radiologists. All the radiologists were trained to use the teleradiological tools on the workstations through tutorials and four teaching sessions. Following the shift, radiologists had to respect a rest period of 1 day.

Radiological Reports

The reports were all created in French. The organization of IMADIS radiological reports is standardized and organized into five separate sections: title, indication, protocol, results and conclusion. For normal examinations, radiologists can use a standardized report for any body region. Otherwise, radiologists fill free-text areas through typing and/or using speech recognition software (Dragon Medical Direct, Nuance Healthcare; Burlington, MA, USA). Spelling mistakes are highlighted in real time to ease manual corrections. After report completion, radiologists must validate it and select if the examination was (i) normal, (ii) pathological in relationship with the symptoms or (iii) pathological for fortuitous/unexpected reasons.

Since most radiologists use standardized reports for normal examinations and radiographs, we applied the following inclusion criteria: (i) examination with at least one pathological finding; (ii) CT or MRI examination; (iii) examination involving only one scanned body region among the brain, spine, head and neck, vasculature, chest, abdomen-pelvis and musculoskeleton; and (iv) examination with the five sections completed (Fig. 1). We also excluded radiographs because they are interpreted separately during dedicated daytime sessions within 24 h following their acquisitions, hence, without the same context of emergency. Examinations involving two or more body regions were excluded as a mean to standardize the size of the reports. Indeed, the numbers of words and sentences in the reports for ≥ 2 body regions would logically expand due to the increase in structures to describe and analyze.

Non-Textual Data Collection

Patients' Data

For each examination, we collected the patient's age, sex and hospital of origin.

Furthermore, we proposed several exploratory non-textual variables capturing different aspects of the generation of the radiological reports in order to understand what would explain their structure and their content:

Radiologists' Data

We reported the gender and experience of the IMADIS radiologist who interpreted the examination. The experience was categorized as senior if a medical degree had been earned at least 2 years ago and junior otherwise (which corresponded to ≥ 7 years of experience in emergency radiology in the French radiological curriculum versus < 7 years). For exploratory purpose, the experience was also investigated continuously as the number of years since medical degree. Additionally, we reported if the professional skills (i.e. field of expertise in radiology, acquired during a 2-year fellowship in a specialized radiology department following medical degree) for senior radiologists matched the scanned body region of the examination. Indeed, because many fields of expertise could be required for the same body region (for instance, interpreting an examination involving the abdominopelvic region could require urology, or gynecology, or vascular or hepatobiliary and digestive radiological competence according to the French radiological curriculum and possible residencies in specialized radiology departments), we added this label only for pediatric, head and neck, brain, musculoskeletal, and chest examinations for further subgroup analysis.

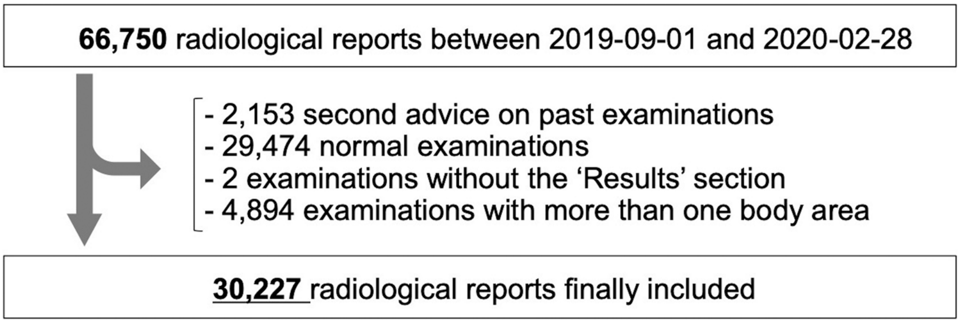
66,750 radiological reports between 2019-09-01 and 2020-02-28
- 2,153 second advice on past examinations
- 29,474 normal examinations
- 2 examinations without the 'Results' section
- 4,894 examinations with more than one body area
30,227 radiological reports finally included

Fig. 1 Study flow chart

Organizational Data

We collected the day and the hour at which the report was completed, which were categorized as days of the week (from Monday to Friday) versus days of the weekend (Saturday and Sunday), and in time slots (8 a.m.–6 p.m., 6 p.m.–0 a.m., 0 a.m.–4 a.m., and 4 a.m.–8 a.m., depending on the main breaks and/or changes of the radiologists' teams). For each examination, we calculated how many pathological examinations were reported by all radiologists from 1 h before to 1 h after and named this variable 'workload', as an estimator of intensification of the activity during busy period. We also computed the number of pathological examinations previously interpreted by the radiologist who reported the current examination during the on-call period, as an estimator of accumulated work during the shift.

Examination-Related Data

We reported the scanned body region and whether it was a pediatric examination (patient's age < 15 years—because patients with age ≥ 15 years are usually referred to adult radiological departments in most French University Hospitals), and we simplified this to 'Type of examination' (categorized as: abdominopelvic, brain, chest, head and neck, musculoskeletal, pediatrics spine and medullary, vascular, and others). We also reported the imaging modality and the emergency level. Indeed, when requesting an examination, the emergency physician has to indicate the priority of the request (i.e. emergency level, categorized as 'extreme emergency', 'common emergency' and 'organizational emergency') to modulate the speed of patient workflow management.

Text Post-Processing

Text post-processing was performed with R (v.4.0.4, Vienna, Austria) on the '18' part of the report using the *stringr*, *stringi* and *tidytext* packages [12]. We focused on the '18' section because it was typically the longest, written in a free-text format with full sentences and punctuation, and was the most likely to be modified based on different contexts.

Enumeration Analysis

In the raw text, we counted (1) the number of question marks and named this variable 'number of questionings'. Indeed, in French radiological reports, it is not uncommon for radiologists to question hypotheses and possibilities separated with question marks, following the depiction of their findings, which can reflect doubt and uncertainty. We also counted (2) the number of negative formulations (for instance, in English: 'neither', 'nor', 'not', 'no', 'absence of', 'without', 'lack of', 'no evidence for/of', … etc.).

The text was then converted to lower case, and we performed the following transformations to obtain the post-processed text:

- French accents, which are present upon some vowels (for instance 'é', 'ê', 'è' or 'ë'), were removed, and the special character 'ç' was changed to 'c' because they are a frequent source of spelling mistakes (the same transformations were performed in the lexicons used in the following paragraphs);
- Punctuations such as '…', '?' and '!' were transformed to '.';
- Indentations were removed;
- Compound words were counted as a single word;
- Commas indicating decimal between two digits of the same number were removed to count this number as a single word;
- Dates were transformed to a single number and counted as a single word;
- Symbols and abbreviations were expanded to their full form (i.e. ' > ' to 'superior'; ' < ' to 'inferior'; '*' and 'x' to 'multiply'; '/' to 'over'; and ' = ' to 'equal');
- Finally, the residual punctuations (',', '`', "'", '- ', ':' and ';') and symbols after this cleaning were removed.

We counted (3) the number of sentences by counting the number of '.' instances in this post-processed text.

The next step consisted in tokenization by parsing the sentences into single units, herein unique words, which corresponds to a 'bag of words' approach. We used the French dictionary of stop words (SW) provided with the *proustr* package to count and remove SWs [13]. SWs are highly frequent words responsible for irrelevant noise (for instance, 'a', 'an', 'and', 'the'…etc.) [6]. Hence, we obtained for each radiological report (4) the total number of words (including SWs), (5) the average number of words per sentence (including SWs), (6) the number of SWs, (7) the number of non-SW words and (8) the number of unique non-SW words (as a marker of lexical diversity).

Sentiment and Uncertainty Analysis

In this approach, the polarity (either positive or negative), the sentiment (herein with a focus on the 'surprise' sentiment) and the degree of uncertainty expressed in the report are estimated by counting the number of words containing a positive, negative, surprised or uncertain tonality. We used three French lexicons to accomplish this task: the 'polarity' and 'score' lexicons provided by the *rfeel* package and a homemade uncertainty lexicon [14, 15]. The polarity lexicon was composed of 5704 words with negative polarity and 8423 words with positive polarity. The score lexicon was composed of 1182 words expressing surprise. The uncertainty lexicon was established in consensus by six senior radiologists and contained 238 words. Each word in the lexicons was post-processed similarly to the word in the radiological reports. Then, for each report, we calculated (9) the polarity score (= number of positive words − number of negative words), (10) the surprise score (= number of surprise words) and (11) the uncertainty score (= number of words expressing uncertainty and approximations).

The variables depicting all the observations are summarized in Fig. 2.

Statistical Analysis

Statistical analyses were also performed with R. All tests were two-tailed. A P value < 0.05 was deemed significant.

Dimensionality Reduction

Correlations among the textual variables were investigated with Spearman rank tests and correlation matrices. We summarized these variables depicting the reports by using principal component analysis (PCA) with the *Factominer* package after centering and scaling. PCA is a widely used method for dimensionality reduction in which the n initial variables of a dataset (named $X_1, X_2, \ldots, X_n$) are projected on k non-collinear principal components (PCs, with $k < p$) that preserve their variance and still contain most of the initial information. By definition, each PC is a linear combination of the n initial variables (i.e. $\text{PC}_\text{i} = \alpha_{i,1} \times X_1 + \alpha_{i,2} \times X_2 + \ldots + \alpha_{i,n} \times X_{n,}$ where $i \in [1, k]$), but the importances of each initial variable in the value of a given PC are not the same. Thus, the initial textual variables ($n = 11$) were synthesized into a lower number of PCs. We decided to select the first 3 PCs only (named PC1, PC2 and PC3) in terms of cumulative percentage of variance in order to facilitate the visualization and understanding of the results, and minimize the numbers of tests. To understand which properties of the radiological reports each of these 3 PCs reflected, we provided their three most important contributors among the initial textual variables.

Univariate Analysis

Univariate associations among the organizational, radiologist-related and examination-related variables and the textual variables, and the first three PCs were evaluated by using Spearman rank tests (for numerical variables), unpaired Wilcoxon tests (for binary categorical variables) or Kruskal–Wallis tests (for categorical variables with > 2 levels). P values were corrected for multiple comparisons using Benjamini–Hochberg corrections. Projections of the

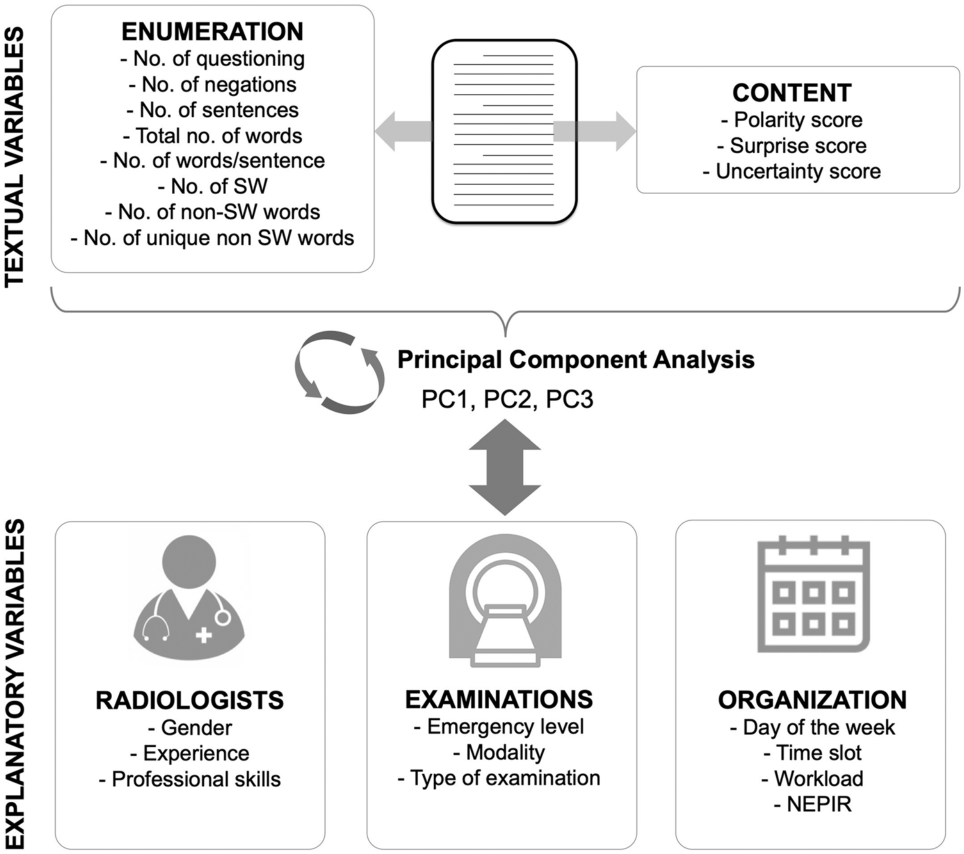

Fig. 2 Assessment of relationships between textual and non-textual variables used in the study. Abbreviations: NEPIR number of pathological examinations previously interpreted by the radiologist, SW stop words

non-textual dichotomized variables on the first three PCs were visualized.

Multivariate Assessment

Multiple-way analyses of variances (mANOVA) were conducted to understand which non-textual variables would influence the content of the radiological reports as synthesized by the first three PCs. Only variables that were correlated with the PCs in univariate analysis were included in this final analysis. Continuous explanatory variables were regularly categorized beforehand.

Results

Study Population (Table 1)

Of the 66,750 reports identified over the study period, 30,227 were ultimately included (Fig. 1). The median age of the patients was 64.8 years (range: 0–105), including 13,984/30,227 (46.3%) women.

There were 183 on-call times with 6 to 23 radiologists on duty per on-call time. The average number of reports per partner hospital was 496 ± 366 over the study time.

Regarding radiologists, there were 50/165 (30.3%) women and 67/165 (40.6%) seniors. On average, the mean number of examinations dictated per radiologists per nightshift (8 p.m.–8 a.m.) was 24 ± 3.4 (median: 23.3, range: 16–34); the mean number of examinations dictated per radiologists per dayshift (8 a.m.–8 p.m.) was 15 ± 5.1 (median: 14, range: 5–41). The radiologists reported an average of 213 ± 162 pathological examinations over the study period (median: 152, range: 4–844), with an average interpretation time of 19.2 ± 11.8 min per pathological examination (median: 18 min, range: 0–111).

Synthesizing the Descriptors of the Radiological Reports with PCA

The summary statistics of the variables describing the structure and the content of the emergency radiological reports are displayed in Table 2.

Using PCA, 93.7% of the variance within the dataset was captured with the six first PCs. The first three PCs represented 71.3% of the variance (47.2%, 14.6% and 9.43%, respectively). The distribution of the potential descriptors along the three main PCs is shown in Fig. 3.

The three main contributors of PC1 were the total number of words (18.3%), non-SW words (18.1%) and unique non-SW words (17.9%), which suggests that PC1 mostly reflected the length of the reports and their lexical diversity.

Table 1 Characteristics of the study population

Characteristics	Radiological reports
Patients	
Age (years)	64.8 (0–105)
Gender	
Women	13,984/30,227 (46.3%)
Men	16,243/30,227 (53.7%)
Organization	
Day of the week	
Monday to Friday	19,573/30,227 (64.8%)
Weekend	10,654/30,227 (35.2%)
Time slot	
8 a.m.–6 p.m	7863/30,227 (26%)
6 p.m.–0 a.m	12,786/30,227 (42.3%)
0 a.m.–4 a.m	6492/30,227 (21.5%)
4 a.m.–8 a.m	3086/30,227 (10.2%)
Workload	22.5 ± 9.6
NEPIR	11.5 ± 8.6
Radiologists	
Interpreted by	
Junior radiologist	14,751/30,227(48.8%)
Senior radiologist	15,476/30,227 (51.2%)
Years of experience since MD	3.4 ± 5.1
Interpreted by	
Woman radiologist	7177/30,227 (23.7%)
Man radiologist	23,050/30,227 (76.3%)
Matching with the radiologist's skills*	
No	14,018/15,143 (92.6%)
Yes	1125/15,143 (7.4%)
Examinations	
Emergency level	
Extreme emergency	1563/30,227 (5.2%)
Common emergency	25,909/30,227 (85.7%)
Organizational emergency	2755/30,227 (9.1%)
Modality	
CT scan	29,716/30,227 (98.3%)
MRI	511/30,227 (1.7%)
Type of examination	
Pediatrics§	545/30,227 (1.8%)
Abdominopelvic	13,160/30,227 (43.5%)
Brain	8538/30,227 (28.2%)
Head and neck	485/30,227 (1.6%)
Musculoskeletal	1422/30,227 (4.7%)
Spine and medullary	1114/30,227 (3.7%)
Chest	4153/30,227 (13.7%)
Vascular	798/30,227 (2.6%)
Other*	12/30,227 (0.04%)

Abbreviations: *NEPIR* number of pathological examinations previously interpreted by the radiologist *Competence matching was assessed for certain body part imaged that corresponded to only one field of expertise, i.e. pediatrics, brain, head and neck, chest and musculoskeletal

§Pediatrics examinations were distributed as follows: abdominopelvic: 114/545 (20.9%), brain: 265/545 (48.6%), head and neck: 89/545 (16.3%), musculoskeletal: 48/545 (8.8%), spine and medullary: 9/545 (1.7%), chest: 28/545 (5.1%) and vascular: 1/545 (0.2%)

Table 2 Summary of the textual variables and contribution to the first three principal components

Characteristics	Mean ± sd	Median (range)	Contribution to PC1	Contribution to PC2	Contribution to PC3
No. of questioning	0.1 ± 0.4	0 (0–6)	0.8	4.6	**29.1**
No. of negative formulation	6.3 ± 3.1	6 (0–23)	7.7	**19.4**	0.1
No. of sentence	10.4 ± 3.7	10 (1–34)	12.2	14.9	2.1
No. of non-SW words	58.2 ± 22.9	55 (7–355)	**18.1**	0.8	0.2
No. of words	96.8 ± 38.6	92 (11–508)	**18.3**	1.1	0.3
No. of SW	38.6 ± 16.7	36 (1–168)	16.6	1.4	0.4
No. of words per sentence	10 ± 5.9	8.9 (2.8–258)	0.3	**38.4**	4
No. of unique non-SW	52.2 ± 18.6	50 (7–256)	**17.9**	0.5	0.2
Polarity scoring	1.7 ± 5.6	2 (-21–38)	0.8	**18.9**	**38.1**
Surprise scoring	3.4 ± 3.4	3 (0–56)	3.5	0	1.9
Uncertainty scoring	1.1 ± 1.2	1 (0–15)	3.9	0	**23.6**

The three most important contributors to each PC are in bold and underlined

Significant correlations are highlighted in bold and underlined

Abbreviations: *no.* number, *PC* principal component, *sd* standard deviation, *SW* stop words

The three main contributors of PC2 were the number of words per sentence (38.4%), the number of negations (19.4%) and polarity (18.9%), which suggests that PC2 mostly reflected a mixture of sentence complexity and negative–positive depictions.

The three main contributors of PC3 were polarity (38.1%), the number of questionings (29.1%) and the number of words expressing uncertainty (23.6%), which suggests that PC3 mostly reflected doubt and ambiguity in addition to length.

Univariate Analysis

Correlations with Organizational Variables

Table 3 shows the correlations among all the explanatory variables and the first three PCs (other correlations with raw text variables are displayed in Supplementary Data 1). Representations of binarized and numeric organizational variables along the three first PCs are given in Fig. 4A, B.

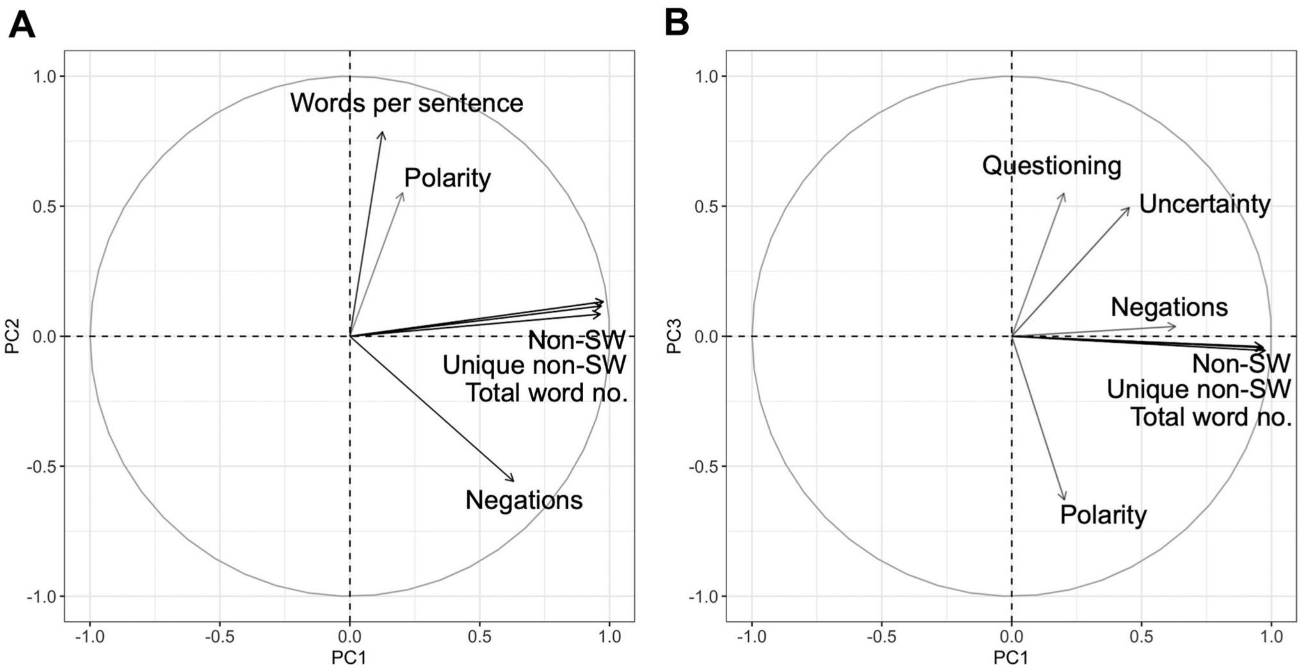

Fig. 3 Synthesis of the most contributive textual variables by using principal component analysis. **A** Representations of the textual variables along the first principal component (PC1, x-axis) and the second principal component (PC2, y-axis). **B** Representations of the textual variables along the first principal component (PC1, x-axis) and the third principal component (PC3, y-axis). Other abbreviations: no. number, SW stop words

Table 3 Univariate correlations between the organizational, radiologist-related and examination-related variables and the first three principal components synthesizing the textual variables

Characteristics	PC1		PC2		PC3	
	Value[a]	*P* value	Value[a]	*P* value	Value[a]	*P* value
Organizational variables						
Day of the week						
Week days	0.08 ± 2.3	**< 0.0001*****	0 ± 1.29	0.3834	0 ± 1.03	0.4378
Weekend	− 0.15 ± 2.23		− 0.01 ± 1.23		− 0.01 ± 1	
Time slot						
8 a.m.–6 p.m	− 0.21 ± 2.3	**< 0.0001*****	0.1 ± 1.27	0.3834	− 0.03 ± 0.98	0.4378
6 p.m.–0 a.m	0.15 ± 2.35		0.02 ± 1.27		− 0.01 ± 1.05	
0 a.m.–4 a.m	− 0.07 ± 2.17		− 0.11 ± 1.28		0.04 ± 1.02	
4 a.m.–8 a.m	0.09 ± 2.09		− 0.12 ± 1.2		0.04 ± 0.99	
Workload						
	0.014 (0.002–0.026)	**0.0210***	− 0.001 (− 0.012–0.01)	0.9055	0.002 (− 0.01–0.013)	0.7924
NEPIR						
	− 0.06 (− 0.072 − − 0.048)	**< 0.0001*****	− 0.073 (− 0.084 − − 0.062)	**< 0.0001*****	0.016 (0.005–0.027)	**0.0073***
Examination-related variables						
Modality						
CT scan	0 ± 2.28	**0.0002*****	− 0.02 ± 1.26	**< 0.0001*****	0.01 ± 1.01	**< 0.0001*****
MRI	0.23 ± 1.95		1.05 ± 1.36		− 0.61 ± 1.11	
Emergency level						
Extreme emergency	1.05 ± 2.66	**< 0.0001*****	0.64 ± 1.53	**< 0.0001*****	− 0.56 ± 1.19	**< 0.0001*****
Usual emergency	0 ± 2.22		− 0.05 ± 1.24		0.03 ± 1.01	
Organizational emergency	− 0.6 ± 2.4		0.14 ± 1.24		0 ± 0.94	
Type of examination						
Abdominopelvic	0.66 ± 2.06	**< 0.0001*****	− 0.56 ± 1.12	**< 0.0001*****	0.4 ± 0.95	**< 0.0001*****
Neuroimaging	− 0.27 ± 2.33		0.53 ± 1.06		− 0.62 ± 0.91	
Head and neck	− 0.43 ± 2.1		0.34 ± 1.15		− 0.12 ± 0.78	
Musculoskeletal	− 2.96 ± 1.54		0.75 ± 1.17		0.17 ± 0.66	
Pediatrics	− 1.02 ± 1.94		0.37 ± 1.39		− 0.12 ± 0.86	
Spine	− 1.23 ± 2.07		0.7 ± 1.25		0.04 ± 0.78	
Chest	− 0.29 ± 1.64		0.02 ± 1.1		0.02 ± 0.95	
Vascular	1.53 ± 3.12		0.72 ± 2.14		− 0.29 ± 1.24	
Other	− 4.87 ± 0.34		2.15 ± 0.48		0.42 ± 0.2	
Radiologist-related variables						
Radiologist's gender						
Woman	0.53 ± 2.48	**< 0.0001*****	0.13 ± 1.57	**< 0.0001*****	0.03 ± 1.15	0.8819
Man	− 0.17 ± 2.19		− 0.04 ± 1.15		− 0.01 ± 0.98	
Radiologist's experience						
Junior	0.05 ± 2.18	**< 0.0001*****	− 0.01 ± 1.3	0.0623	0.03 ± 1.06	**0.0072***
Senior	− 0.05 ± 2.37		0.01 ± 1.23		− 0.02 ± 0.98	
Radiologist's skill[b]						
No match	− 0.51 ± 2.19	**< 0.0001*****	0.39 ± 1.13	**0.0016***	− 0.35 ± 0.95	**< 0.0001*****
Matching	− 1.18 ± 2.49		0.45 ± 0.97		− 0.18 ± 0.97	

Abbreviations: *NEPIR* number of pathological examinations previously interpreted by the radiologist, *PC* principal component

[*] $P < 0.05$; **$P < 0.005$; ***$P < 0.001$. All P values were corrected for multiple testing. Significant correlations are highlighted in bold and light gray

[a] Value is Spearman correlation coefficient with 95% confidence interval for numerical explanatory variables (NEPIR and workload) or mean with standard deviation otherwise

[b] Competence was assessed on a subsample of the whole cohort for which the body part imaged could only correspond to one field of expertise, i.e. pediatrics, brain, head and neck, chest and musculoskeletal

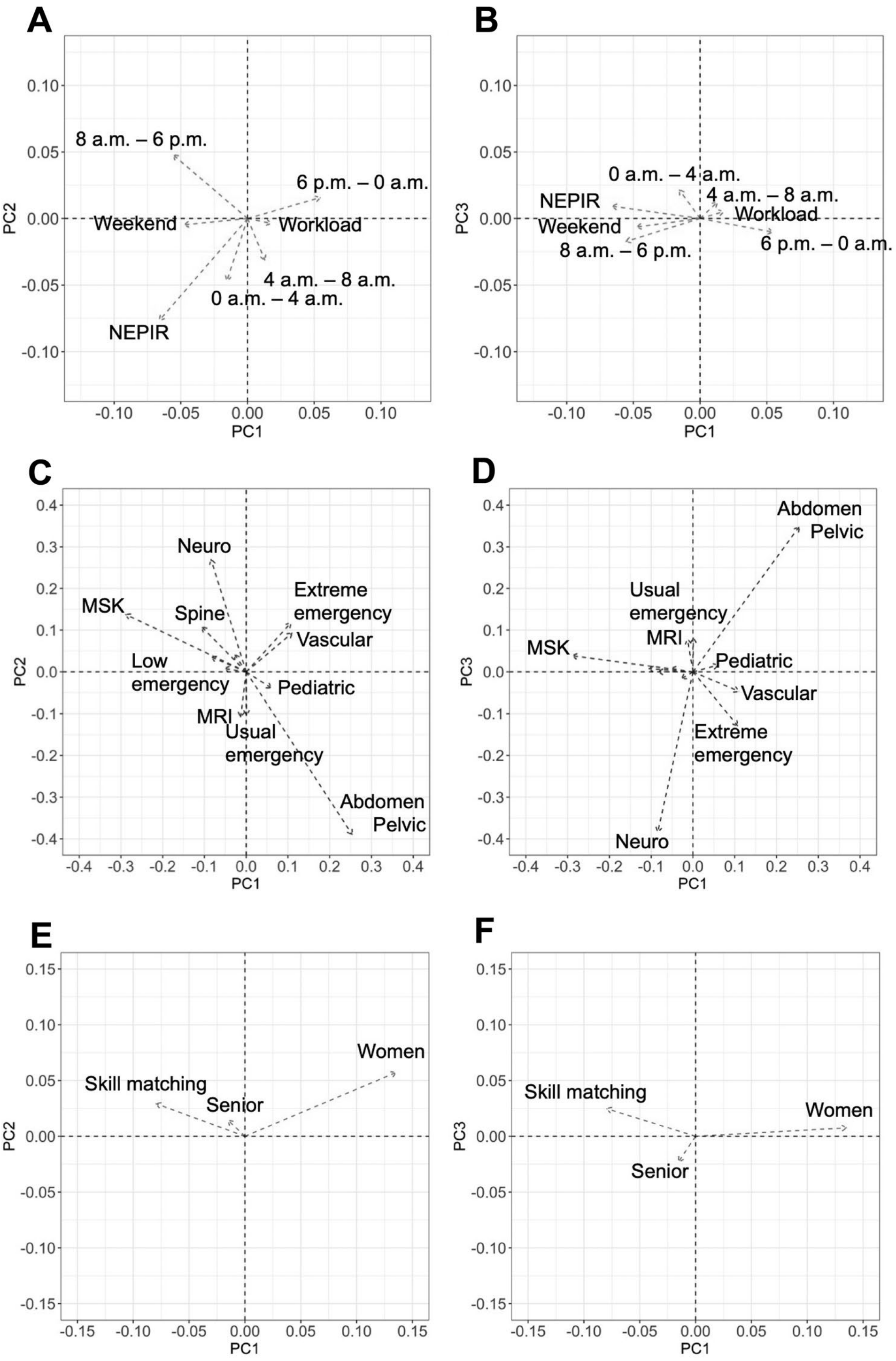

Fig. 4 Representations of the explanatory variables along the first three principal components (PCs). Projections of organizational variables depending on PC1 and PC2 **A**, and PC1 and PC3 **B**. Projections of examination-related variables depending on PC1 and PC2 **C**, and PC1 and PC3 **D**. Projections of radiologist-related variables depending on PC1 and PC2 **E**, and PC1 and PC3 **F**. Other abbreviations: MSK musculoskeletal examination, NEPIR number of pathological examinations previously interpreted by the radiologist

PC1 (i.e. length and lexical diversity) was significantly lower during the weekend, during the 8 a.m.–6 p.m. and 0 a.m.–4 a.m. periods, and when the number of examination previously interpreted by the radiologist increased (all P values < 0.0001). A weak but significant positive correlation was found between PC1 and workload ($P = 0.0210$). PC2 (i.e. sentence complexity and positive–negative depictions) was significantly lower when the number of examination previously interpreted by the radiologist increased ($P < 0.0001$). PC3 (i.e. doubt and ambiguity) was significantly positively correlated with the number of examination previously interpreted by the radiologist ($P = 0.0073$).

Correlations with Examination-Related Variables

The three examination-related variables (namely: type of examination, imaging modality and emergency level) were all strongly associated with PC1, PC2 and PC3 (P value range: < 0.0001–0.0002) (Table 3, Fig. 4D).

In detail, compared with CT-scan reports, we found that MRI reports were longer overall (average total number of words = 103.8 ± 33.9 versus 96.7 ± 38.6 for CT scan, $P < 0.0001$), with more words per sentence (12.5 ± 8.5 versus 9.9 ± 5.8 for CT scan, $P < 0.0001$) and fewer negative formulations (4.97 ± 2.38 versus 6.32 ± 3.08 for CT scan, $P < 0.0001$) (Supplementary Data 1).

Regarding the emergency level, the 'extreme emergency' label led to significantly longer reports (total number of words = 116.8 ± 45.5 versus 95.72 ± 37.87 for others, $P < 0.0001$) with a wider vocabulary (number of unique non-SW words = 61.9 ± 21.4 versus 51.68 ± 18.28 for others, $P < 0.0001$) but a lower uncertainty (0.89 ± 1.16 versus 1.1 ± 1.16 for others, $P < 0.0001$) and a similar number of negative formulations (6.3 ± 3.28 versus 6.3 ± 3.06 for others, $P = 0.9928$).

Regarding the type of examination, the longest reports were observed for abdominopelvic and vascular examinations (total number of words = 105.4 ± 35.5 and 127.4 ± 56.4, respectively), while the shortest were musculoskeletal examinations (54 ± 27.5 words). Abdominopelvic examinations also displayed the largest number of negative formulations (7.68 ± 2.69 negations versus 5.24 ± 2.92 for others, $P < 0.0001$). Chest examinations demonstrated the most important number of words explaining uncertainty (1.64 ± 1.08 versus 1 ± 1.15 for others, $P < 0.0001$).

Correlations with Radiologist-Related Variables

The value of PC1 (i.e. length and lexical diversity) was significantly higher with female radiologists compared with male radiologists (0.53 ± 2.48 versus − 0.17 ± 2.19, $P < 0.0001$), junior radiologists compared with senior radiologists (0.05 ± 2.18 versus − 0.05 ± 2.37, $P < 0.0001$) and radiologists whose radiological skill did not match the examination (− 0.51 ± − 1.18 versus − 1.18 ± 2.49, $P < 0.0001$) (Table 3, Fig. 4E, F). For instances, the total number of words was 106.2 ± 42.9 on average for women radiologists versus 93.9 ± 36.6 for men, 97.4 ± 36.5 for junior radiologists versus 96.2 ± 40.4 for senior (spearman rho = − 0.058 [95% confidence interval: − 0.069; − 0.046, $P < 0.0001$] for years of experience since medical degree) and 89.9 ± 37.2 for radiologists without skill matching versus 79.9 ± 41.6 with, all P values < 0.0001 (Supplementary Data 1).

The value of PC2 (i.e. sentence complexity and positive–negative depictions) was also significantly higher in female radiologists (0.13 ± 1.57 versus − 0.04 ± 1.15, $P < 0.0001$) and in radiologists with skill matching (0.45 ± 0.97 versus 0.39 ± 1.13, $P = 0.0016$).

PC3 (i.e. doubt and ambiguity) was similar between male and female radiologists ($P = 0.8819$) but slightly significantly lower with senior radiologists (− 0.02 ± 0.98 versus 0.03 ± 1.06, $P = 0.0072$) and without skill matching (− 0.35 ± 0.95 versus − 0.18 ± 0.97, $P < 0.0001$).

Multivariate Analysis (Table 4)

Figure 5 synthesizes all the correlations found among the explanatory variables, the report-related variables and the synthetic PCs.

Since the assessment of skill matching was achieved for 15,143 patients, mANOVAs were also performed on this subcohort without any missing values (Table 4). The results for the entire cohort without considering skill matching are given in Supplementary Data 2.

Overall, PC1 (i.e. length and lexical diversity) was significantly influenced by the day of the week, workload, time slot, number of examination previously interpreted by the radiologist, type of examination, emergency level and radiologist gender (range of P values < 0.0001–0.0029). Sixteen interactions were significant, of which the most influential were the radiologist's gender and experience (F value = 86.1, $P < 0.0001$).

PC2 (i.e. sentence complexity and positive–negative depictions) was significantly influenced by the number of examination previously interpreted by the radiologist, the type of examination, the emergency level, the imaging modality and the radiologist's experience (range of P values < 0.0001–0.0032). Two interactions were significant: the type of examination with the level of emergency ($P = 0.0046$) and with the radiologist's skill ($P = 0.0117$).

PC3 (i.e. doubt and ambiguity) was significantly influenced by the type of examination, the emergency level and the interaction between the emergency level and the radiologist's skill ($P < 0.0001$, < 0.0001 and 0.0204, respectively).

Discussion

While many clinical, epidemiological and managerial applications can be expected from the application of NLP to emergency radiological reports, the factors influencing the means of expression of radiologists have been poorly investigated thus far. Yet, identifying such influential factors could help us to improve the language used in radiological reports, their quality and to homogenize our practices. Overall, our findings show that the length, structure and content of emergency radiological reports were significantly different depending on organizational, radiologist- and examination-related characteristics.

We used an original but mandatory method to simplify the numerous variables depicting the emergency radiological reports, namely dimensionality reduction with PCA. PCA is widely used in radiomics and machine-learning approaches for instance to summarize numerous collinear numeric variables into synthetic non-collinear fewer ones, named PCs. The drawback of PCA is that the meaning of the 3 identified PCs is not obvious because they are synthetic combinations of the 11 initial text variables. This explains why we first investigated what our main PCs reflected and estimated. It should be noted that other methods could have been used to synthesize the text-related variables, such as variable clustering (with k-means or hierarchical clustering), but PCA is generally the first method tried and enables simple projections of new supplementary variables on the variables graph, which is helpful to visualize similarity and correlations [16].

First, we observed that two examination-related variables significantly influenced the main three PCs in the multivariate assessment: the type of examination and the emergency level. Vascular and abdominopelvic examinations provided the longest reports, which was expected given the numerous structures to describe leading to larger amounts of negative formulations. By contrast, musculoskeletal examinations comprised the shortest reports, with immediate and direct depiction of the pathological findings without exhaustive depictions of normal findings. Interestingly, chest CTs contained the highest number of words expressing uncertainty, which is in agreement with the often non-specific semiology of parenchymal findings. Additionally, high-priority examinations led to longer reports, with a more diverse vocabulary and longer sentences, together with a distinct polarity, fewer negative formulations and fewer words expressing uncertainty. This result could be partly counterintuitive because the findings of extreme emergency examinations must be transmitted as quickly as possible to the emergency physician in stressful conditions. We explain this difference by the radiologists' need to be as clear as possible with the emergency physician in this setting to avoid missing the transmission of life-threatening information. Furthermore, as part of the IMADIS workflow, the radiologist interpreting such examination must contractually phone the emergency physicians within 15 min after receiving the images. Thus, exchanges between the radiologist and the requiring physician could lead to answers to additional precise questions that could change the therapeutic management or reinsurance. Interestingly, Hassanpour et al. also identified the marked influence of examination-related variables on radiological reports [17]. The authors performed an unsupervised k-means clustering of the reports in free-text format and found that topics were mostly classified according to the anatomic site and the imaging modality.

Regarding the radiologists' characteristics, the multivariate analysis emphasized the significant influence of gender on PC1 related to the length of the reports and their lexical

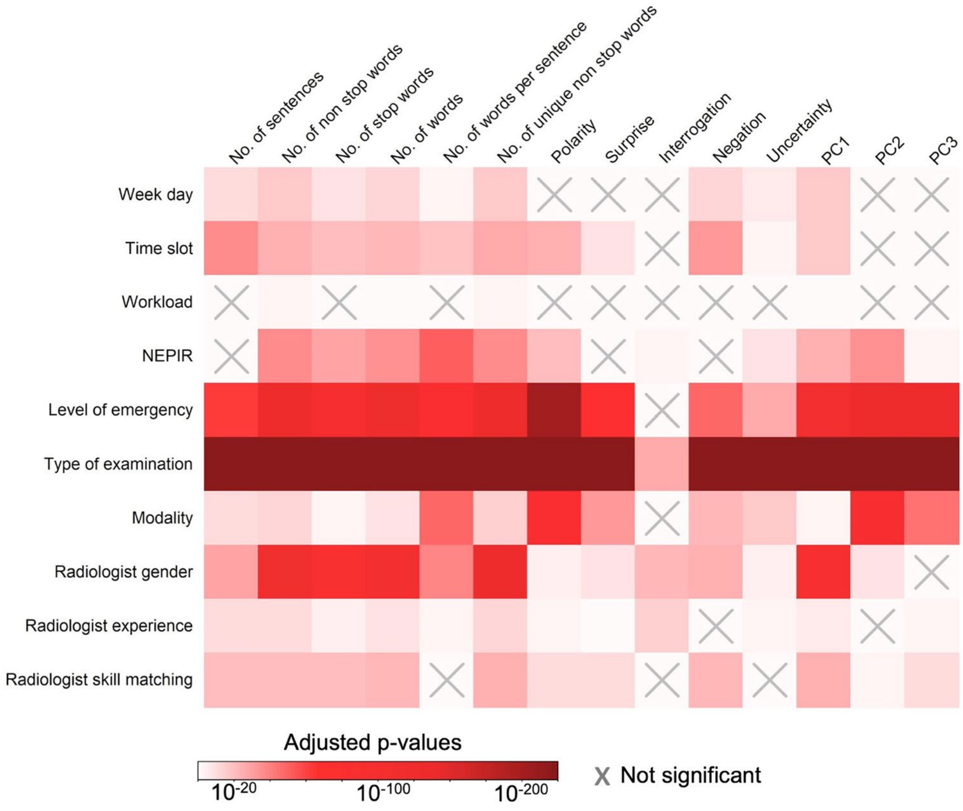

Fig. 5 Correlation matrix between textual variables (columns) and explanatory variables (rows). Non-significant results are indicated with a gray cross. Abbreviations: NEPIR number of pathological examinations previously interpreted by the radiologist, no. number, PC principal component, SW stop words

Table 4 Multivariate correlations between the organizational, radiologist-related and examination-related variables and the first three principal components synthesizing the textual variables in the subcohort of 15,143 patients with radiologist skill assessment

Characteristics	***F* value**	***P* value**
mANOVA for PC1		
Day of the week	78.9	< 0.0001***
Workload[a]	4	0.0029**
Time slot	21.4	< 0.0001***
NEPIR[b]	11.2	< 0.0001***
Type of examination	575.3	< 0.0001***
Emergency level	245	< 0.0001***
Modality	0.1	0.7607
Radiologist’s experience	1.2	0.2792
Radiologist’s gender	249.3	< 0.0001***
Radiologist’s skill	0	0.8911
Day of the week × Workload	3	0.016*
Day of the week × Time slot	5.6	0.0008*
Day of the week × NEPIR	2.6	0.032*
Day of the week × Type of examination	3.1	0.0141*
Day of the week × Radiologist's gender	26	< 0.0001***
Time slot × Radiologist's experience	3.6	0.0136*
Time slot × Radiologist's gender	4.6	0.0033**
Type of examination × Emergency level	3.1	0.0017**
Type of examination × Radiologist's gender	7.1	< 0.0001***
Emergency level × Modality	21.8	< 0.0001***
Emergency level × Radiologist's experience	3.9	0.0212*
Emergency level × Radiologist's gender	3	0.0483*
Radiologist's experience × Radiologist's gender	86.1	< 0.0001***
Type of examination × Radiologist's competence	9.1	< 0.0001***
Modality × Radiologist's skill	5.1	0.0245*
Radiologist's gender × Radiologist's skill	7.7	0.0054*
mANOVA for PC2		
NEPIR	14.7	< 0.0001***
Type of examination	192.1	< 0.0001***
Emergency level	17.6	< 0.0001***
Modality	49.3	< 0.0001***
Radiologist's experience	8.7	0.0032**
Radiologist's skill	3.1	**0.078**
Type of examination × Emergency level	2.8	0.0046**
Type of examination × Radiologist's skill	3.2	0.0117*
mANOVA for PC3		
NEPIR	1.5	0.2143
Type of examination	513.7	< 0.0001***
Emergency level	40.6	< 0.0001***
Modality	1.8	0.1739
Radiologist's experience	1.3	0.2584
Radiologist's skill	0.1	0.7418
Emergency level × Radiologist's skill	3.9	0.0204*

Competence was assessed on a subsample of the whole cohort for which the body part imaged could only correspond to one field of expertise, i.e. pediatrics, brain, head and neck, chest and musculoskeletal

Abbreviations: *mANOVA* multiple-way analysis of variance, *NEPIR* number of pathological examinations previously interpreted by the radiologist, *PC* principal component

$^{*}P<0.05$; $^{**}P<0.005$; $^{***}P<0.001$. Significant correlations are highlighted in bold. Only significant second-order correlations (x) are shown

[a] Workload was categorized as 0–10, 11–20, 21–30, 31–40 and > 40

[b] NEPIR was categorized as 0–10, 11–20, 21–30, 31–40, 41–50 and > 50

diversity, but not on the two other PCs. Indeed, radiologists' experience was associated with PC2, while their skill area in interaction with the priority of the examination was associated with PC3. The correlation between gender and length of application letters to residency programs in various medical fields has already been stressed in several studies [18–21]. Here, we observed a similar trend in medical documents produced by physicians for physicians.

Regarding organization variables, we proposed two complementary variables: workload, which aims at representing intensification of the activity during short window of time, and number of examinations previously interpreted by radiologist, which aims at representing the accumulation of work over the shift. We hypothesized that these variables would correlate with stress and tension, and tiredness and weariness, respectively. Interestingly, the organizational variables did not have a strong influence on the textual PCs. Their projection on the first three PCs was less pronounced than the projections of examination- and radiologist-related variables. Although weak, the multivariate assessment emphasized correlations with PC1 but with smaller *F* values compared to other variable categories.

Correlations among workload, end of the shift (especially after 10 h or within the last 4 h) and decreased diagnostic accuracy in on-call radiologists have been well described by Hanna et al. [22, 23]. However, the authors did not investigate tiredness and stress as intermediate links among workload, late hours and accumulation of work, with resulting discrepancies. Indeed, real-time monitoring of stress and tiredness is particularly difficult during on-call duty. Questionnaires could be biased by their inherent retrospective nature, and devices recording physical manifestations could hamper medical practice [24, 25].

Herein, the organization of the on-call period was different from classical organizations because of several measures to limit tiredness and weariness [26–29]. The environment surrounding radiologists working at IMADIS was designed by architects specializing in air traffic control centers and emergency regulation centers. Ergonomics of the chairs, tables, lighting, sound absorption and workstations are optimized in all three teleradiological interpretation centers. Regular break times are planned, and separated common and individual rest rooms are provided. Radiologists work in small groups, thereby enabling constant interactions and stimulation. Finally, phone disturbance and interpretation disruption for protocoling examinations are strongly limited by a rotation system during which a selected senior radiologist is fully dedicated to these two tasks every 2 h. Consequently, we hypothesize that these measures reduced the impact of organizational variables on textual summary variables.

Alternative methods could have been used for text mining. We purposely chose classical methods in this exploratory hypothesis-generating study, i.e. text cleaning, unigram tokenization, SW removal and matching with handcrafted and public French lexicons, and PCA for dimensionality reduction. The inclusion period was focused on the 6 months preceding the beginning of the COVID-19 pandemic in mainland France because emergency radiological activity has been deeply modified afterwards with a lower volume of examinations and a strong increase in chest CT scans at the expense of other body regions [30, 31]. We studied the results section because it is typically the longest and most 'literary' section, whereas the conclusion summarizes the main results as short, bulleted points. We excluded normal examinations and multiple-body region examinations to limit bias. Indeed, prewritten normal reports are largely used in the absence of pathological findings, and each textual variable would have increased with the increase in scanned body regions, resulting in confusion regarding the influence of examination-related variables. We did not use lemmatization, n-gram analysis or more complex metrics (such as term frequency-inverse document frequency, type-token ratio, Yule I or first-order Markov models) to deepen the report characterization so as to limit the number of variables to simple and explainable ones, but NLP-based models are generally enhanced by using these approaches [32–34].

Moreover, looking at the emergency radiological reports achieved by 165 radiologists points out that the important variability of the radiologists' written expression to convey their analysis of CT scans or MRIs. Previous studies have shown that using structured radiological reports (i.e. a homogeneous and ordered method to express the results) enables a better adhesion to guidelines and is generally preferred by clinicians because they provide a better transmission of pathological findings in shorter interpretation delays [35–40]. It can also be hypothesized that structured and standardized reports could reduce the inter-radiologist variability. However, there is a lack of data regarding the clinical impact on patients' care of structured reports compared with unstructured reports.

Interestingly, since this study, IMADIS has implemented several structured and standardized reports in the interpretation workflow for pathological examinations (and notably chest CT scans for suspicion of COVID-19 infection), by using pre-filled fields (related to the disease semiology, its severity and its complications) that radiologists can complete with pre-specified options and conclude with published international classifications.

Finally, the transposability of our results (based on French language) to the English language could be questioned. We believe that the principle of the methods is completely applicable to any other language but the exact replicability of our results cannot be claimed because of the possible influence of the language itself and also because of the possible cultural, societal or educational differences between countries (even for different countries using the same language). Only international linguistic studies based on radiological reports could investigate these hypotheses.

Our study has limitations. First, except for uncertainty, the French lexicons used were not dedicated to medical language. Second, it should be noted that words expressing uncertainty were not weighted differently depending on the degree of uncertainty. Furthermore, a more comprehensive and subtle quantification of uncertainty could have been reached using sentence tokenization instead of a bag-of-words approach. For instance, herein, 'this may represent' was equivalent to 'this almost certainly represents' in terms of uncertainty expression, although the first sentence conveys much more uncertainty. Third, the teams were mostly composed of young radiologists (< 45 years old), whereas older people are more sensitive to alterations of the circadian cycle, suggesting potential interactions between age and time slot with the textual variables [41]. Fourth, our dataset did not include the number of diagnostic errors, a tiredness assessment nor an evaluation of stress and tension during the shift, although investigating correlations with textual variables would have been helpful to develop original warning tools for emergency radiologists. Thus, although we hypothesized that workload was associated with stress and tension, and number of previously interpreted examinations was associated with tiredness and weariness, we could not affirm these associations. Furthermore, it can be argued that we did not include the normal examinations in our calculation of these two organization variables. In fact, the proportion of normal examinations over time remained globally constant (between 41 and 48% whatever the hour of interpretation of time slot) and since they can be rapidly managed thanks to pre-existing standardized and structured reports, we do not believe that it would have significantly biased our results. Fifth, other categorizations for the hour of interpretation could have been proposed, but we purposely chose to categorize it according to main breaks and changes in the radiological teams. Sixth, although we deliberately focused on pathological examinations to avoid the bias due to the use of pre-existing standardized reports for normal examination, we could not be certain that some radiologists did insert these pre-existing reports and just changed some sentences related to the pathological findings. Seventh, we did not assess the variability in radiologists' written expression for the same examinations, i.e. our data were not paired. Finally, other methods could have been proposed to estimate the workload, for instance: interpretation within the last 4 h versus before, or within 10 ± 2 h of night work has been previously used [22, 23].

Conclusion

To conclude, our study highlights correlations between the structure and contents of medical texts produced by emergency radiologists during on-call duty and several organizational, examination- and radiologist-related features. We believe that these findings emphasize the subjectivity and variability in the way radiologists express themselves during their clinical activity and, consequently, stress the need to consider these influential features when developing NLP-based models.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10278-022-00619-6.

Declarations

Ethics Approval This research study was conducted retrospectively from data obtained for clinical purposes. We consulted extensively with the IRB of the French Ethics Committee for the Research in Medical Imaging (CERIM, IRB-approval number: CRM-2106–168) who determined that our study did not need ethical approval. An IRB official waiver of ethical approval was granted from the IRB of CERIM.

Conflict of Interest The authors declare no competing interests.

References

1. Lakhani, P., Kim, W., Langlotz, C.P.: Automated detection of critical results in radiology reports. J Digit Imaging. 25, 30–36 (2012). https://doi.org/10.1007/s10278-011-9426-6.
2. Lakhani, P., Kim, W., Langlotz, C.P.: Automated extraction of critical test values and communications from unstructured radiology reports: an analysis of 9.3 million reports from 1990 to 2011. Radiology. 265, 809–818 (2012). https://doi.org/10.1148/radiol.12112438.
3. Wheater, E., Mair, G., Sudlow, C., Alex, B., Grover, C., Whiteley, W.: A validated natural language processing algorithm for brain imaging phenotypes from radiology reports in UK electronic health records. BMC Med Inform Decis Mak. 19, 184 (2019). https://doi.org/10.1186/s12911-019-0908-7.
4. Pons, E., Braun, L.M.M., Hunink, M.G.M., Kors, J.A.: Natural language processing in radiology: a systematic review. Radiology. 279, 329–343 (2016). https://doi.org/10.1148/radiol.16142770.
5. Cai, T., Giannopoulos, A.A., Yu, S., Kelil, T., Ripley, B., Kumamaru, K.K., Rybicki, F.J., Mitsouras, D.: Natural language processing technologies in radiology research and clinical applications. Radiographics. 36, 176–191 (2016). https://doi.org/10.1148/rg.2016150080.
6. Chen, P.-H.: Essential elements of natural language processing: what the radiologist should know. Acad Radiol. 27, 6–12 (2020). https://doi.org/10.1016/j.acra.2019.08.010.
7. Les urgences hospitalières, miroir des dysfonctionnements de notre système de santé, http://www.senat.fr/rap/r16-685/r16-6852.html, last accessed 2021/07/14.
8. Selvarajan, S.K., Levin, D.C., Parker, L.: The increasing use of emergency department imaging in the United States: is it appropriate? AJR Am J Roentgenol. 213, W180–W184 (2019). https://doi.org/10.2214/AJR.19.21386.
9. Bruno, M.A., Duncan, J.R., Bierhals, A.J., Tappouni, R.: Overnight resident versus 24-hour attending radiologist coverage in academic medical centers. Radiology. 289, 809–813 (2018). https://doi.org/10.1148/radiol.2018180690.
10. Kalyanpur, A.: The role of teleradiology in emergency radiology provision. Radiol Manage. 36, 46–49 (2014).

11. Crombé, A., Lecomte, J.-C., Banaste, N., Tazarourte, K., Seux, M., Nivet, H., Thomson, V., Gorincour, G.: Emergency teleradiological activity is an epidemiological estimator and predictor of the covid-19 pandemic in mainland France. Insights Imaging. 12, 103 (2021). https://doi.org/10.1186/s13244-021-01040-3.
12. Silge, J., Robinson, D.: Text mining with R: a tidy approach. O'Reilly Media, Inc. (2017).
13. Fay, C.: ColinFay/proustr. (2021).
14. Fay, C.: ColinFay/rfeel. (2020).
15. Abdaoui, A., Azé, J., Bringay, S., Poncelet, P.: FEEL: a French expanded emotion lexicon. Lang Resources & Evaluation. 51, 833–855 (2017). https://doi.org/10.1007/s10579-016-9364-5.
16. Chavent, M., Kuentz-Simonet, V., Liquet, B., Saracco, J.: ClustOfVar: an R package for the clustering of variables. Journal of Statistical Software. 50, 1–16 (2012). https://doi.org/10.18637/jss.v050.i13.
17. Hassanpour, S., Langlotz, C.P.: Unsupervised topic modeling in a large free text radiology report repository. J Digit Imaging. 29, 59–62 (2016). https://doi.org/10.1007/s10278-015-9823-3.
18. Turrentine, F.E., Dreisbach, C.N., St Ivany, A.R., Hanks, J.B., Schroen, A.T.: Influence of gender on surgical residency applicants' recommendation letters. J Am Coll Surg. 228, 356-365.e3 (2019). https://doi.org/10.1016/j.jamcollsurg.2018.12.020.
19. Kobayashi, A.N., Sterling, R.S., Tackett, S.A., Chee, B.W., Laporte, D.M., Humbyrd, C.J.: Are there gender-based differences in language in letters of recommendation to an orthopaedic surgery residency program? Clin Orthop Relat Res. 478, 1400–1408 (2020). https://doi.org/10.1097/CORR.0000000000001053.
20. Lin, F., Oh, S.K., Gordon, L.K., Pineles, S.L., Rosenberg, J.B., Tsui, I.: Gender-based differences in letters of recommendation written for ophthalmology residency applicants. BMC Med Educ. 19, 476 (2019). https://doi.org/10.1186/s12909-019-1910-6.
21. Powers, A., Gerull, K.M., Rothman, R., Klein, S.A., Wright, R.W., Dy, C.J.: Race- and gender-based differences in descriptions of applicants in the letters of recommendation for orthopaedic surgery residency. JB JS Open Access. 5, (2020). https://doi.org/10.2106/JBJS.OA.20.00023.
22. Hanna, T.N., Loehfelm, T., Khosa, F., Rohatgi, S., Johnson, J.-O.: Overnight shift work: factors contributing to diagnostic discrepancies. Emerg Radiol. 23, 41–47 (2016). https://doi.org/10.1007/s10140-015-1355-0.
23. Hanna, T.N., Lamoureux, C., Krupinski, E.A., Weber, S., Johnson, J.-O.: Effect of shift, schedule, and volume on interpretive accuracy: a retrospective analysis of 2.9 million radiologic examinations. Radiology. 287, 205–212 (2018). https://doi.org/10.1148/radiol.2017170555.
24. Whitehead, L.: The measurement of fatigue in chronic illness: a systematic review of unidimensional and multidimensional fatigue measures. Journal of Pain and Symptom Management. 37, 107–128 (2009). https://doi.org/10.1016/j.jpainsymman.2007.08.019.
25. Patterson, P.D., Weaver, M.D., Fabio, A., Teasley, E.M., Renn, M.L., Curtis, B.R., Matthews, M.E., Kroemer, A.J., Xun, X., Bizhanova, Z., Weiss, P.M., Sequeira, D.J., Coppler, P.J., Lang, E.S., Higgins, J.S.: Reliability and validity of survey instruments to measure work-related fatigue in the emergency medical services setting: a systematic review. Prehosp Emerg Care. 22, 17–27 (2018). https://doi.org/10.1080/10903127.2017.1376134.
26. Boiselle, P.M., Levine, D., Horwich, P.J., Barbaras, L., Siegal, D., Shillue, K., Affeln, D.: Repetitive stress symptoms in radiology: prevalence and response to ergonomic interventions. Journal of the American College of Radiology. 5, 919–923 (2008). https://doi.org/10.1016/j.jacr.2008.01.014.
27. Rohatgi, S., Hanna, T.N., Sliker, C.W., Abbott, R.M., Nicola, R.: After-hours radiology: challenges and strategies for the radiologist. AJR Am J Roentgenol. 205, 956–961 (2015). https://doi.org/10.2214/AJR.15.14605.
28. Prabhu, S.P., Gandhi, S., Goddard, P.R.: Ergonomics of digital imaging. Br J Radiol. 78, 582–586 (2005). https://doi.org/10.1259/bjr/51650129.
29. Harisinghani, M.G., Blake, M.A., Saksena, M., Hahn, P.F., Gervais, D., Zalis, M., da Silva Dias Fernandes, L., Mueller, P.R.: Importance and effects of altered workplace ergonomics in modern radiology suites. Radiographics. 24, 615–627 (2004). https://doi.org/10.1148/rg.242035089.
30. Vatele, J., Gentile, S., Thomson, V., Devictor, B., Cloux, M., Girouin, N., Bratan, F., Bergerot, J.-F., Seux, M., Banaste, N., Tazarourte, K., Gorincour, G.: Teleradiology as a relevant indicator of the impact of COVID-19 pandemic management on emergency room activities: a nationwide worrisome survey. Insights Imaging. 12, 30 (2021). https://doi.org/10.1186/s13244-021-00964-0.
31. Gibson, A.L., Chen, B.Y., Rosen, M.P., Paez, S.N., Lo, H.S.: Impact of the COVID-19 pandemic on emergency department CT for suspected diverticulitis. Emerg Radiol. 27, 773–780 (2020). https://doi.org/10.1007/s10140-020-01865-3.
32. Yule, G.U.: On sentence-length as a statistical characteristic of style in prose: with application to two cases of disputed authorship. Biometrika. 30, 363–390 (1939). https://doi.org/10.2307/2332655.
33. Zech, J., Pain, M., Titano, J., Badgeley, M., Schefflein, J., Su, A., Costa, A., Bederson, J., Lehar, J., Oermann, E.K.: Natural language-based machine learning models for the annotation of clinical radiology reports. Radiology. 287, 570–580 (2018). https://doi.org/10.1148/radiol.2018171093.
34. Montemurro, M.A., Zanette, D.H.: Universal entropy of word ordering across linguistic families. PLOS ONE. 6, e19875 (2011). https://doi.org/10.1371/journal.pone.0019875.
35. Barbosa, F., Maciel, L.M.Z., Vieira, E.M., Azevedo Marques, P.M. de, Elias, J., Muglia, V.F.: Radiological reports: a comparison between the transmission efficiency of information in free text and in structured reports. Clinics (Sao Paulo). 65, 15–21 (2010). https://doi.org/10.1590/S1807-59322010000100004.
36. Larson, D.B., Towbin, A.J., Pryor, R.M., Donnelly, L.F.: Improving consistency in radiology reporting through the use of department-wide standardized structured reporting. Radiology. 267, 240–250 (2013). https://doi.org/10.1148/radiol.12121502.
37. Powell, D.K., Silberzweig, J.E.: State of structured reporting in radiology, a survey. Acad Radiol. 22, 226–233 (2015). https://doi.org/10.1016/j.acra.2014.08.014.
38. Segrelles, J.D., Medina, R., Blanquer, I., Martí-Bonmatí, L.: Increasing the efficiency on producing radiology reports for breast cancer diagnosis by means of structured reports. A comparative study. Methods Inf Med. 56, 248–260 (2017). https://doi.org/10.3414/ME16-01-0091.
39. Griffin, A.S., Mitsky, J., Rawal, U., Bronner, A.J., Tessler, F.N., Hoang, J.K.: Improved quality of thyroid ultrasound reports after implementation of the ACR thyroid imaging reporting and data system nodule lexicon and risk stratification system. J Am Coll Radiol. 15, 743–748 (2018). https://doi.org/10.1016/j.jacr.2018.01.024.
40. Stanzione, A., Ponsiglione, A., Cuocolo, R., Rumolo, M., Santarsiere, M., Scotto, R., Viceconte, G., Imbriaco, M., Maurea, S., Camera, L., Gentile, I., Brunetti, A.: Chest CT in COVID-19 patients: structured vs conventional reporting. Eur J Radiol. 138, 109621 (2021). https://doi.org/10.1016/j.ejrad.2021.109621.
41. Turek, F.W., Penev, P., Zhang, Y., van Reeth, O., Zee, P.: Effects of age on the circadian system. Neurosci Biobehav Rev. 19, 53–58 (1995). https://doi.org/10.1016/0149-7634(94)00030-5.