English for Specific Purposes 47 (2017) 1–14

Contents lists available at ScienceDirect

English for Specific Purposes

journal homepage: http://ees.elsevier.com/esp/default.asp

Profiling maritime communication by non-native speakers: A quantitative comparison between the baseline and standard marine communication phraseology

Peter John [a,b,*], Benjamin Brooks [b], Ulf Schriever [b]

[a] Jade University of Applied Sciences, Faculty of Maritimes Studies, Weserstr. 52, D-26931 Elsfleth, Germany
[b] Australian Maritime College, University of Tasmania, Maritime Way, Newnham TAS 7248, Australia

A R T I C L E I N F O

Article history:
Available online 12 April 2017

Keywords:
Bridge team communication
Maritime English
Genre analysis
Relative word frequencies
Maritime key words

A B S T R A C T

This paper compares ESP communication by non-native speakers of Maritime English with communication outside a nautical setting in order to profile its structural idiosyncrasy. Vocabulary growth, word frequencies, lexical and key word densities, and grammar diversity as dependent linguistic variables observed in transcribed full-mission simulation exercises are contrasted to the Brown Corpus, the Vienna-Oxford International Corpus of English and the Standard Marine Communication Phrases (SMCP). Using quantitative linguistics, inherent structural patterns of nautical team communication are identified and similarities and variations highlighted. Significant differences found in all linguistic features are gauged by means of the Probability of Superiority (PS) effect size. A *linguistic profile* is created which quantifies the observed language patterns and provides a quantitative model for the linguistic genre of this particular discourse community. The model fills the gap of quantitative research on empirical bridge team communication samples and delivers a valid tool for estimating the magnitude of observed linguistic effects.
© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Communication on board ships has long been identified as a decisive factor for safe navigation. This importance becomes especially evident whenever a ship accident occurs, as in the disastrous evacuation procedures of passenger ships Costa Concordia and Sewol, to state two recent examples. Research has found that communication problems alone cause almost half of all marine accidents whilst miscommunication is a contributory factor in nearly all shipping accidents (for an overview, cf. John, Brooks, Wand, & Schriever, 2013; Möckel, Brenker, & Strohschneider, 2014). Most communication on board ships is verbal, but although Voyage Data Recorders (VDR) are installed on modern ships to record, amongst other information, all utterances made by navigational officers, a very limited number of authentic bridge team communication samples is available for linguistic research. This notorious scarcity of empirical information has been highlighted by Dževerdanović-Pejović (2013).

Linguistic research relies on observations on how people communicate. However, given the limited scope of authentic speech samples little quantitative research has been conducted in the domain of the bridge team communication

* Corresponding author. Weserstr. 52, D-26931 Elsfleth, Germany.
E-mail address: peter.john@jade-hs.de (P. John).

http://dx.doi.org/10.1016/j.esp.2017.03.002
0889-4906/© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

discourse community in order to determine which speech patterns are actually used by seafarers to assess situations, carry out navigational tasks and avoid dangerous situations, and to what extent these speech acts differ from spontaneous verbal communication outside the maritime world (Cole & Trenkner, 2012; Pritchard, 2003; Trenkner, 1996; Weeks, 1997).

While it is true that authentic communication from a ship's bridge is not available in an annotated corpus for quantitative research purposes, future nautical officers are trained in full-mission ship handling simulators which replicate the navigational tasks carried out on board ship. The research presented in this paper makes use of these simulation exercises to overcome the scarcity of available on-board speech by using audio-recorded maritime simulation sessions. A verbatim transcript of the recorded communication allows corpus linguistics techniques to be used with the aim to discriminate idiosyncratic language patterns of seafarers.

1.1. Bridge team communication

Bridge team communication is a generic term for spontaneous speech acts by nautical officers who navigate the ship as a team. It shares most of the characteristics of team communication outside a nautical setting. However, given the very specific work environment in which it takes place, bridge team communication also differs regarding the team's composition, communication channels and the scope of its content. Teams on board sea-going ships are nearly always multinational and multicultural, with a substantial number of nationalities and ethnicities working very closely together and sharing one environment in which they work and live (cf. Deboo, 2004; Horck, 2005; Noble, Vangehuchten, & Van Parys, 2011). On merchant vessels, bridge teams are usually composed of the captain or shipmaster, the first, second and third navigational officer and a helmsman. Crew members work in shifts covering the ship operation twenty-four hours a day, seven days a week (cf. Jensen et al., 2006). Depending on the ship type and deployment area work shifts of four or six hours are customary. Due to this organisational scheme two to three people usually work together during their shift after which the team composition changes. One officer, who need not be the highest ranked, has the Command of Navigation (CoN) and is therefore responsible for all decisions made and actions taken during each watch.

While bridge team members engage in direct, face-to-face communication in order to assess situations and make decisions, they also communicate via UHF radio with other crew members located in different areas of the ship, e.g. in the engine room, on deck, etc., which extends the bridge team to a distributed team. Communication is also undertaken by VHF radio with the shore-based Vessel Traffic Service (VTS), tugs and other ships, and sometimes via satellite or mobile phone with the shipping company, the charterer's agents and other people ashore, so that a virtual team environment is created (for an overview on virtual team communication cf. Potter & Balthazard, 2002). If no other common language is available, the International Maritime Organisation (IMO) stipulates that crew members shall communicate in English as a lingua franca.

Given the importance of bridge team communication for the safe operation of a ship, the scarcity of publications of quantitative research in this specific discourse domain is rather surprising.

1.2. Research question and hypotheses

Observing bridge team communication in full-mission simulation opens up possibilities to gain an insight into the structure of naturally occurring language in a unique English for Specific Purposes (ESP) environment. Different speech patterns can be analysed and inferences made on their effectiveness in given situations. By contrasting maritime with non-maritime communication, similarities and differences can be singled out, and the appropriateness or idiomaticity of the language use is discernible.

The adopted research approach aims to contribute to a quantitative model of the language variety or *genre* of Bridge Team Communication as a *sub-genre* (cf. Baker & Ellece, 2011, p53) of Maritime English. By analysing a series of linguistic variables it sets out to identify and define this specific ESP variety using a descriptive approach.

These objectives lead to the following research question:

To what extent do the speech patterns of bridge team communication by non-native speakers of English in full-mission simulation differ lexically and grammatically from other, non-nautical communication?

In order to answer this general research question the following null hypotheses are formulated:

$H_0$1: The inter-textual vocabulary growth does not differ significantly between bridge team and other, non-nautical communication.

$H_0$2: The relative word frequency distribution does not differ significantly between bridge team and other, non-nautical communication.

$H_0$3: The distribution of content words does not differ significantly between bridge team and other, non-nautical communication.

$H_0$4: The distribution of nautical key words does not differ significantly between bridge team and other, non-nautical communication.

$H_0$5: The part-of-speech diversity distribution observed in bridge team communication does not differ significantly from other, non-nautical verbal communication.

The five hypotheses aim to profile the bridge team members' overall speech behaviour and to quantify linguistic variables which can be assumed to be idiosyncratic for the given ESP environment. $H_0$1 compares the number of different words (types) that can be expected for a given total word count (tokens). $H_0$2 studies how these word types are distributed. $H_0$3 computes differences in the corpora's lexical density. $H_0$4 identifies the distribution of specific maritime key words, and $H_0$5 highlights differences in grammar diversity.

2. Data sampling

This research compares verbatim transcripts of full-mission bridge team simulation with three different text corpora: the Brown Corpus of Standard American English (1979, 1961), the Vienna-Oxford International Corpus of English (VOICE 2013) and the SMCP text collection developed by the authors based on the Standard Marine Communication Phrases (International Maritime Organization, 2001).

2.1. Text corpora

The Brown Corpus of Standard American English (henceforth referred to as Brown Corpus) dates back to 1961 and comprises 500 texts, each of which consists of about 2,000 words. The corpus is divided into 15 different categories and contains more than one million words. In spite of it being one of the first big text corpora, it is still intensively used for linguistic research, with over five thousand publications citing it over the past ten years according to a Google Scholar search for "Brown Corpus". Given its extended use as a reference text corpus over many years, this corpus is especially apt for being used in baseline calculations. For this research, the tagged version (form C) of the revised and amplified version of 1979 has been used (Francis & Kucera, 1979).

The Vienna-Oxford International Corpus of English (henceforth referred to as Vienna Corpus) comprises in excess of one million words of naturally occurring face-to-face communication by 1,250 second language speakers with approximately 50 different first languages who use English as a lingua franca in a variety of different speech acts. For this research, the VOICE POS XML 2.0 version has been used. In order to avoid any distortions in the computed linguistic values, the tags for Breathing (BR), Laughter (LA) and Pause (PA) have been removed as these are not included in the other three corpora. The Vienna Corpus reflects a natural usage of English as produced by non-native speakers (Seidlhofer et al., 2013).

The Brown Corpus and the Vienna Corpus are used to compute the baseline communication patterns that can be expected outside the nautical environment of bridge team communication.

The Standard Marine Communication Phrases (SMCP) were introduced by the International Maritime Organization in the year 2001 as a set of "precise, simple and unambiguous" phrases (International Maritime Organization, 2001, p3). These phrases attempt to cover all internal and external communicative situations on board sea-going ships thus reducing "problems of communication [which] may cause misunderstandings leading to dangers to the vessel, the people on board and the environment" (International Maritime Organization, 2001, p3) by using "a simplified version of maritime English in order to reduce grammatical, lexical and idiomatic varieties to a tolerable minimum, using standardized structures for the sake of its function aspects" (International Maritime Organization, 2001, p12). The "ability to use and understand the IMO SMCP is required for the certification of officers in charge of a navigational watch on ships of 500 gross tonnage or more" (International Maritime Organization, 2001, p3) and made compulsory under the International Convention on Standards of Training, Certification and Watchkeeping (STCW) for Seafarers, 1978, as revised in 1995 (International Maritime Organization, 1995). In 2010, the Manila Diplomatic Conference on the STCW Convention further strengthened the importance of effective oral communication (cf. Trenkner & Cole, 2010).

In line with its intended use, the SMCP have been published as a manual to be used in education and training. All phrases are displayed in a compacted form by using wildcard characters. In its current form the SMCP can therefore not be used for quantitative text analysis. For this reason, they have been re-worded by the authors into discrete sentences as shown in Table 1. Further examples can be found at www.smcpexamples.com (Gregorič & John, 2013).

According to McEnery and Wilson (2001, p29) "any collection of more than one text can be called a corpus [...], hence a corpus may be defined as any body of text. It need imply nothing more. But the term 'corpus' when used in the context of modern linguistics tends most frequently to have more specific connotations than this simple definition provides for". Given

Table 1
Original and re-worded SMCP phrases.

Original SMCP phrase	Re-worded SMCP phrase
A1/1.1.7.1: I am / MV … ∼ not under command. ∼ adrift. ∼ drifting at ... knots to ... (cardinal points/half cardinal points). ∼ drifting into danger.	I am not under command. Motor vessel Pi not under command. I am adrift. I am drifting at two knots to North North West. I am drifting into danger. Motor vessel Pi adrift. Motor vessel Pi drifting at two knots to North North West. Motor vessel Pi drifting into danger.

that the re-worded SMCP do not constitute a *collection of more than one text*, they are henceforth not referred to as a text *corpus* but as the SMCP text *collection*.

The SMCP text collection can be considered the prescriptive language standard against which all bridge team communication can be analysed structurally and lexically. It constitutes the highest level of idiomaticity and is therefore used in this research as the reference for lexico-grammatical frequency and pattern analysis.

2.2. Bridge team transcript

The verbatim transcript of bridge team communication is based on observational data obtained from training exercises recorded in the years 2013 and 2014 at the Maritime Faculty of Jade University of Applied Sciences in Germany. The exercises included 10 bridge teams involving a total of 23 under-graduate students in their final year of Nautical Sciences who volunteered to participate in the exercises. No participants withdrew from their participation, so no attrition effects must be assumed. All students were German nationals and non-native speakers of English out of which 21 stated their mother tongue as German, one as German and Dutch and one as Tagalog. On a five-point Likert scale (excellent, very good, good, satisfactory, poor), five students rated their English skills as very good, 14 as good and four as satisfactory (median = 3 “good”). They had worked on board sea-going ships for at least one year (median = 13 months). The sample included 22 male students and one female. Each exercise was recorded over 60 min which leads to a total recording time of 600 min. The data were collected with the informed consent of all participants involved and in compliance with the Social Sciences Human Research Ethics regulations of Jade University of Applied Sciences (Germany) and of the University of Tasmania (Australia). The transcripts were made by the first author and validated by the co-authors. Ambiguous or unintelligible words were marked with a wildcard character.

The recorded bridge team communication includes typical standard tasks carried out by navigational officers including route planning, being underway (proceeding) and assessing possible risks to navigation. Participating students communicated face-to-face with the members of their bridge team and by VHF radio with the simulated Vessel Traffic Service (VTS), the Maritime Rescue Co-ordination Centre (MRCC) and with other simulated ships. They also used UHF radio to talk to their own ship’s bosun. All radio communication partners were senior navigation officers working at the simulation facilities except for the other ships which were equipped with volunteering students. Transcript excerpt 1 includes some typical elements of direct, face-to-face communication among bridge team members.

Transcript excerpt 1

speaker	**utterance**
1) shipmaster	What is the next course?
2) officer	Next course in the channel is two two one.
3) shipmaster	Two two one.
4) officer	Uhum.
5) officer	We are now almost abeam, erm, this one.
6) officer	But the vessel is really hard to steer because she is rather short and the current strong and every time is, every time going like, like that.
7) shipmaster	But on the other way it is, er, she’s very high manoeuvrable.
8) officer	Yes, we have six zero rudder angle, yeah, what do you expect?
9) shipmaster	But if you, but if you use six zero it will be really difficult to...
10) officer	Oh, we really don’t have to use six for the moment.
11) shipmaster	She’s coming very, very quick.
12) officer	Yeah, the course is two two one.
13) shipmaster	Hm.
14) shipmaster	But now I have really to take a close look because there is one vessel coming here.
15) officer	But if you look, have a look here.
16) shipmaster	Yeah, oh, this one.
17) shipmaster	Yeah, she’s turning very good, so I would turn then really hard.
18) officer	Yeah, from this buoy to this buoy the channel is a straight line, so.
19) shipmaster	Okay, yeah, no problem.
20) officer	Yeah.

Transcript excerpt 1 clearly illustrates how the shipmaster and nautical officer discuss a possible risk, develop a shared mental model and agree on measures to be taken. The excerpt includes typical elements of verbal communication and some elements of its specific ESP context.

3. Data analysis

In the analysis of the collected data the following steps are undertaken for accepting or rejecting each null hypothesis: Firstly, the reasoning for the hypothesis and descriptive statistical information are presented for the linguistic feature to be analysed. Secondly, the analytical method is presented and pre-requisites are defined and tested. Thirdly, the analysis is

Table 2
Overview of word count and ratios in analysed text corpora.

	Baseline		Empirical data	Prescriptive reference
	Vienna Corpus	Brown Corpus	Bridge team transcript	SMCP text collection
Word tokens	1,016,399	1,036,125	43,019	46,529
Word types	17,449	40,187	1,843	1,883
	Ratio 0.017	Ratio 0.039	Ratio 0.043	Ratio 0.040
Content words	394,230	558,066	20,767	31,879
	Ratio 0.387	Ratio 0.538	Ratio 0.482	Ratio 0.685
Adjectives	49,535	71,994	1,559	2,426
	Ratio 0.048	Ratio 0.069	Ratio 0.036	Ratio 0.052
Adverbs	79,155	37,898	4,911	1,183
	Ratio 0.077	Ratio 0.036	Ratio 0.114	Ratio 0.025
Nouns	155,897	270,978	7,592	17,846
	Ratio 0.153	Ratio 0.261	Ratio 0.176	Ratio 0.383
Verbs	92,619	115,842	3,819	4,878
	Ratio 0.091	Ratio 0.111	Ratio 0.088	Ratio 0.104
Numerals	61,354	17,024	2,886	5,546
	Ratio 0.060	Ratio 0.016	Ratio 0.067	Ratio 0.119
Function words	622,169	478,059	22,252	14,650
	Ratio 0.613	Ratio 0.462	Ratio 0.518	Ratio 0.315
Key words	189,397	208,633	13,212	25,601
	Ratio 0.186	Ratio 0.020	Ratio 0.307	Ratio 0.550

carried out, and fourthly, findings are summarised. Table 2 displays word counts for the Brown and Vienna Corpus, the Bridge Team transcript and the SMCP text collection. Given the substantial differences in the number of words each text corpus contains, it also states ratios for the given subcategories, e.g. the ratio of different word classes to the total word count.

Hypothesis $H_0$1. The inter-textual vocabulary growth does not differ significantly between bridge team and other, non-nautical communication.

Vocabulary growth describes the changing relation of word types (vocabulary size) to word tokens (total word count) over an increasing text length (inter-textual growth) or a defined time frame. It has been extensively used for estimating lexical diversity in first and second language learners (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Laufer, 1998; Nagy & Scott, 2000; Verhoeven, van Leeuwe, & Vermeer, 2011) where it is assumed that an increase in vocabulary growth correlates with an advancing language learning process.

In maritime communication, a higher lexical diversity does not necessarily lead to an improved communication by its discourse community. To the contrary, the standard phraseology was introduced as a coded language which reduces lexical richness on purpose to remove any ambiguities and provide a simple and clear language. Table 2 hints at differences between the type-token ratios (TTR) with figures ranging from 0.017 for the Vienna Corpus, 0.039 for the Brown Corpus, 0.043 for the Bridge Team transcript and 0.040 for the SMCP text collection. However, these figures are biased as the TTR was found to differ in relation to the length of the chosen text samples (cf. Covington & McFall, 2010). For this reason, the corpora were split into samples of 43,019 words which is the exact length of the Bridge Team transcript. Following this method, the Brown and Vienna Corpus were both divided into 23 samples totalling 989,437 words in each corpus. In both corpora, the remaining words were disregarded. To take full advantage of the much smaller SMCP text collection, 10 random samples of 43,019 words each were drawn up without removing the extracted samples.

Figure 1 displays the number of text types for the text token chunks in the three corpora and the SMCP text collection.

The SMCP text collection presents a type-token ratio which is much closer to that of the Bridge Team transcript, with TTR values of 0.040 and 0.043, respectively. The expected number of types for any given number of tokens (up to a value of 43,019) can be calculated by the power functions $V(n)_{SMCP} = 0.99n^{1.44}$ (SMCP, $R^2 = 0.989$) and $V(n)_{BT} = 0.02n^{1.91}$ (Bridge Team transcript, $R^2 = 0.995$).

Although it seems obvious that the distribution of the type-token ratio differs clearly, at least between the Bridge Team transcript and the Brown and Vienna Corpus, the four corpora were also tested statistically in order to quantify the significance of the findings, i.e. the probability for the effects to occur by chance.

To be able to compare the samples' variances, these have to be tested first for a normal distribution as this determines which statistical methods can be employed. Anderson–Darling's test was performed for testing the corpora's distribution against a normal distribution, leading to $p = 0.25$ for the Brown Corpus, $p = 0.19$ for the Vienna Corpus, $p = 0.37$ for the SMCP text collection and $p = 0.12$ for the Bridge Team transcript. All samples are therefore assumed to be normally distributed.

The text corpora were also tested for homoscedasticity employing Levene's test for homogeneity of variances. For the four corpora together it resulted in $p < 0.000$ so that a significant difference in the samples' distribution was assumed. Post-hoc comparisons between the Bridge Team transcript lead to $p < 0.000$ for the Brown and the Vienna Corpus and $p = 0.76$ for the SMCP text collection so that a homoscedastic distribution can only be assumed between the Bridge Team transcript and the SMCP text collection.

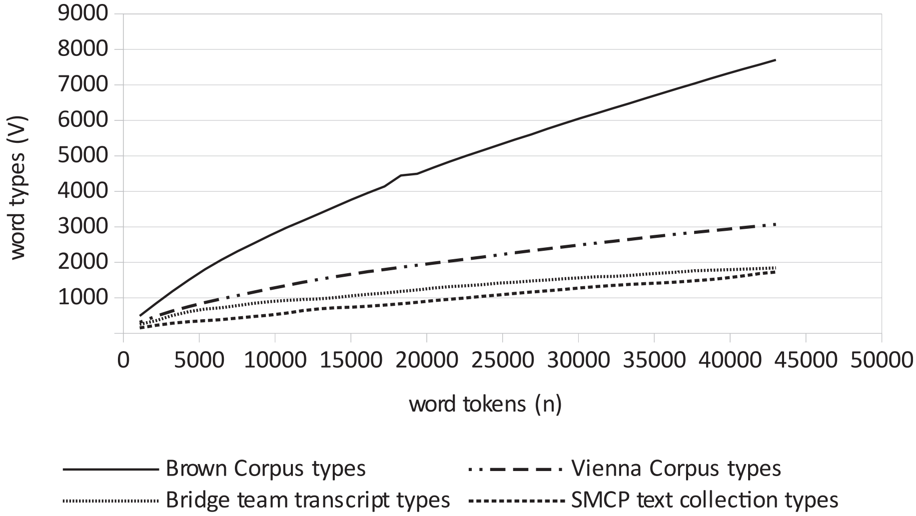

Figure 1. Inter-textual vocabulary growth in text corpora for chunks of 43,019 words.

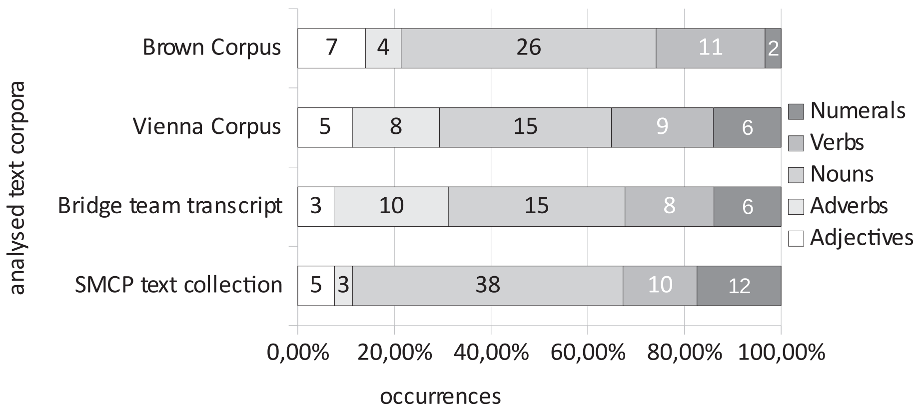

Figure 2. Occurrences of content word classes in text corpora.

Given the text corpora's heteroscedasticity, non-parametric statistical testing methods were chosen. As Figure 1 clearly illustrates, the different word type samples (i.e. the different text corpora) do not originate from the same distribution. The non-parametric Kruskal–Wallis analysis of variance carried out on the four text corpora corroborates these findings with $p < 0.000$ (two-tailed, $\alpha = 0.01$, confidence level 99%). Mann–Whitney U tests were performed post-hoc between the Bridge Team transcript and the other three text corpora. For the SMCP text collection it resulted in $p = 0.005$ and for the Brown and the Vienna Corpus it resulted in $p < 0.000$ (two-tailed, $\alpha = 0.01$, confidence level 99%). The null hypothesis $H_0 1$ must be rejected as the differences in the inter-textual vocabulary growth are highly significant between the text corpora observed.[1]

The vocabulary growth observed in the Bridge Team communication sample differs very significantly from the Brown and Vienna Corpus. Bridge team communication by non-native speakers uses a far more restricted vocabulary size than that observed in written English and, more importantly, than in verbal communication by non-native speakers outside a maritime setting. In comparison with the Brown Corpus, non-native bridge team members can be expected to use between 40% and 50% of word types for any text length of up to 5,000 words, 30–39% for a text length of less than 12,000 words and 29–24% for a text length of up to 43,000 words. For the Vienna Corpus the percentages are 70–80% in texts of less than 5,000 words and 60–79% in texts of less than 43,000 words.

Compared with the SMCP text collection, the Bridge Team communication sample contains only 1–1.9% more types for any given text size up to 43,000 words.

[1] The stated statistical methods were used for all hypothesis testing, given that no two distributions were found to be normally distributed nor were their distributions homoscedastic.

Table 3
Log likelihood statistics for the analysed text corpora.

Word type	LL Bridge Team transcript – Vienna Corpus	LL Bridge Team transcript – Brown Corpus	LL Bridge Team transcript – SMCP text collection
Ahead	+1.7	−86.1	−86.1
Alarm	+24.6	−47.8	−46.7
Anchor	+7.5	−19.6	+11.5
Bridge	+0.7	−30.9	−34.3
Depth	0.0	−29.2	−27.0
Lead	+11.4	−6.7	−7.4
Message	+2.4	−4.2	−2.1
Navigation	+3.8	−17.4	−14.1
Same	+8.7	+37.0	+73.8
Speak	+0.8	+0.3	+65.0
Within	+116.8	+10.9	+13.3

Hypothesis $H_0$2. The relative word frequency distribution does not differ significantly between bridge team members and other, non-nautical communication.

Word frequency distributions provide information on the lexical proximity of texts. Word types that appear at a similar frequency indicate a relatedness of two texts. On the other hand, types that appear more frequently in one text than in another can be considered key words of a specific field covered by the given text.

The frequency profiling technique detects differences of relative word frequencies between any two texts. It is independent from a text's length, so that the Brown and Vienna Corpus could be used without splitting them into smaller fragments. A frequency list of the 406 word types that occurred in all of the four text corpora was produced and the log likelihood statistic (LL) was calculated using a contingency table as suggested by Rayson and Garside (2000). Subsequently, the expected frequency values and the log likelihood statistic itself were computed for each word type. Table 3 presents a selection of the calculated values. The plus sign describes a higher frequency and the minus sign denotes a lower frequency of the word type in the respective text corpus compared to the Bridge Team transcript.

An LL value of zero means that a given word type appears in both text corpora at an identical frequency, and the higher the LL value, the more significant are the differences in the relative frequencies of the two text corpora. Adopting a significance level of $\alpha = 0.01$, the null hypothesis $H_0$2 must be rejected for LL values equal or higher than the critical value of 6.63. Following this procedure for all 406 word types it was found that in comparison with the SMCP text collection, 224 word types presented a significantly different frequency. For the Brown Corpus this value amounted to 305 word types and for the Vienna Corpus to 292. Given the fact that out of a total of 1,843 word types observed in the Bridge Team communication only 406 types were tested for their relative frequencies in the other three text corpora while the remaining 1,383 word types do not appear at all in at least one of the other corpora, $H_0$2 is rejected.

In his paper "Language is never, ever, ever, random", Kilgarriff (2005, p263) points out that "[l]anguage users never choose words randomly, and language is essentially non-random". This non-randomness of language is clearly reflected in relative word frequencies where a log likelihood statistic of zero for all word types is only possible in two identical texts. However, when studying the LL values beyond the mere hypothesis testing, they are indicative of the divergence between the different text corpora. The median LL value for the 406 words shared by the Bridge Team communication and the Brown Corpus is 18.92 while for the Vienna Corpus it is 15.09. As in the case of inter-textual vocabulary growth, the relative word frequencies are more closely related to the utterances by non-native speakers included in the Vienna Corpus than to the more formal English written by native speakers as reflected in the Brown Corpus.

The LL statistic for the comparison with the SMCP text collection is 8.73, a value which is expectedly much closer to the Bridge Team communication than the other, more general text corpora. Here, 44% of the compared word types occurred at a similar relative frequency as opposed to only 25% in the case of the Brown Corpus and 28% in the Vienna Corpus.

Hypothesis $H_0$3. The distribution of content words does not differ significantly between bridge team and other, non-nautical communication.

Words can be differentiated according to their grammatical and lexical function, leading to the lexical density concept which considers the ratio of content words to the total word count (cf. Halliday, 1987). Content words include nouns, verbs, adjectives, adverbs and numerals whereas function words include all other word classes. By computing the lexical density of a given text, inferences can be made about its information content. This is important as communication by work teams mostly aims at interchanging information. The more information is exchanged effectively (i.e. correctly transmitted, received and understood), the higher the linguistic effectiveness of speech acts. In a safety-critical environment like a ship's bridge, this can be a decisive factor. As an example, in transcript excerpt 2 all content words are underlined and the lexical density (LD) is given for each utterance.

Transcript excerpt 2

speaker	utterance	LD
21) shipmaster	I will keep it now with fifty percent and, er.	0.40
22) shipmaster	When we cross all these vessels, then, I'm happy.	0.40
23) officer	But no more traffic is out there.	0.57
24) shipmaster	We really pass, er?	0.50
25) officer	Like this, what's, what's that on the left side?	0.18
26) shipmaster	There's one more vessel, I see already.	0.50
27) shipmaster	Okay, anyway, we'll check this with the radar.	0.44
28) shipmaster	Ha, ha, ha, uh, okay, now I see the buoys.	0.40
29) shipmaster	This is very nice, I can mark them already.	0.44
30) shipmaster	There is one more vessel but it shouldn't be a problem.	0.25
31) shipmaster	Okay, now I will go starboard again, yeah?	0.63
32) officer	Yeah.	0.00

This example illustrates the substantial differences in lexical density. The highest value observed in (31) clearly contains more information than (25) or (32). The word count and ratios for content and function words in the text corpora are presented in Table 2. Figure 2 compares the different content word classes observed.

In the following, the probability for the differences displayed in Figure 2 to occur randomly is computed for all content words together and for each individual content word class. If the calculation was simply performed on the counted words of a particular class in each utterance, this would obviously not take into account the differences in the utterance length (word count per each utterance). For this reason the calculated ratio of a particular word class is considered for any given utterance. So, if an utterance with a total length of ten words contains four content words, its content word ratio of 0.40 is used. If the same utterance contains two adjectives, its adjective ratio of 0.20 is used for the computation.

As Figure 3 well illustrates, the different content word samples (i.e. the different text corpora) do not originate from the same distribution. The non-parametric Kruskal–Wallis analysis of variance carried out on the four text corpora corroborates these findings with $p < 0.000$ (two-tailed, $\alpha = 0.01$). Mann–Whitney *U* tests were performed post-hoc between the Bridge Team transcript and the other three text corpora. For the SMCP and the Vienna Corpus it resulted in $p < 0.000$, and for the Brown Corpus in $p = 0.001$ (two-tailed, $\alpha = 0.01$).

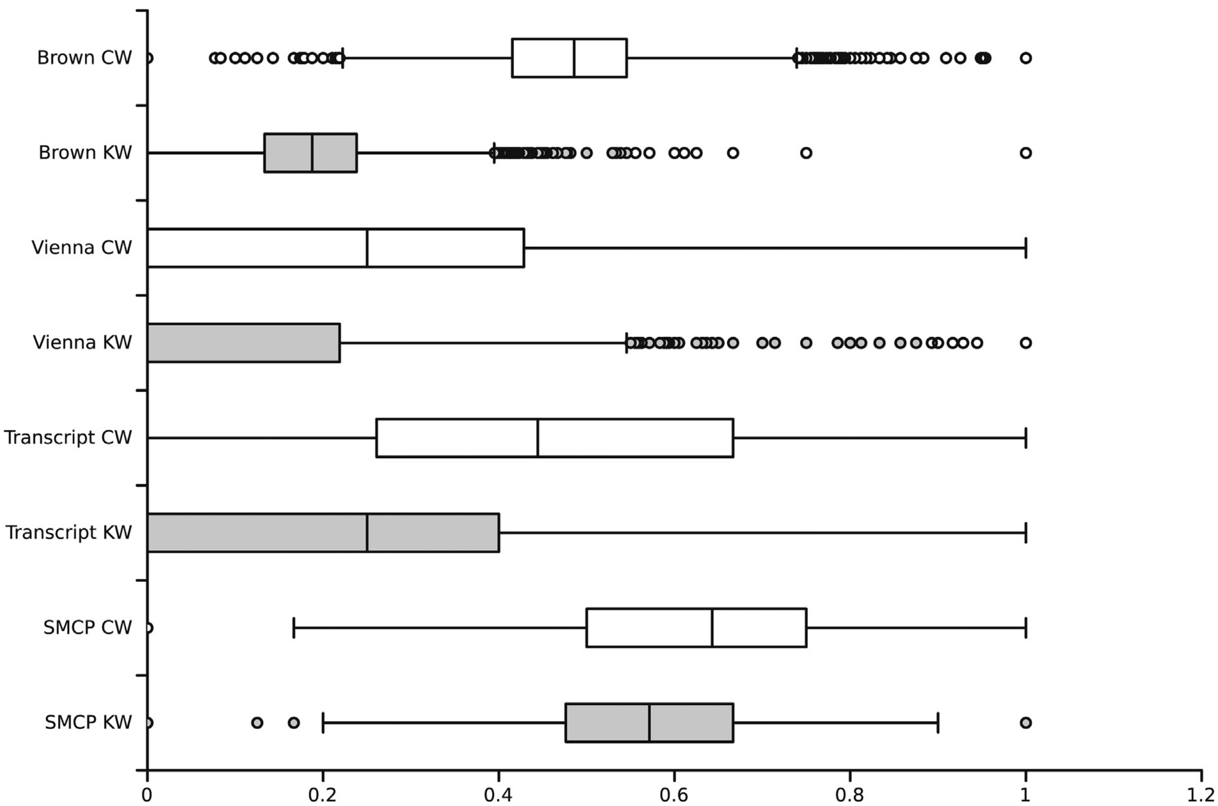

Figure 3. Content word and key word distribution in analysed text corpora.

Table 4
Effect sizes of content word classes.

PS	Content words	Adjectives	Adverbs	Nouns	Verbs	Numerals
Brown Corpus	0.49	0.28	0.54	0.32	0.33	0.44
Vienna Corpus	0.70	0.49	0.61	0.59	0.54	0.58
SMCP text collection	0.30	0.41	0.65	0.18	0.34	0.46

The Kruskal–Wallis test carried out for occurrences of verbs, nouns, adjectives, adverbs and numerals in the different text corpora leads to $p < 0.000$, and post-hoc Mann–Whitney *U* tests on the five word classes between the individual corpora also resulted in $p < 0.000$ for all combinations. In all tests, the significance level was set to $\alpha = 0.01$.

In order to estimate the strength of the differences between the individual content word ratios, the effect size is computed as their Probability of Superiority (PS) which does neither assume a normal nor a homoscedastic distribution of the observed variables. Grissom and Kim (2005, p98) have described this effect size as "the probability that a randomly sampled member of population *a* will have a score (Ya) that is higher than the score (Yb) attained by a randomly sampled member of population *b*."

The effect sizes between the Bridge Team transcript and the other corpora are displayed in Table 4.

$H_0 3$ must be rejected because the distribution of content and function words in bridge team communication differs significantly from other, non-nautical communication. Bridge team communication presents a higher lexical density than the non-nautical, verbal communication collected in the Vienna Corpus (VC). The distribution between both corpora differs strongly with PS = 0.70. The lexical density of bridge teams is nearly identical with the written texts in the Brown Corpus (BC) with PS = 0.49. The medians in both text corpora are relatively close (transcript = 0.44, BC = 0.50 as opposed to VC = 0.25). However, the transcript's lexical density varies more markedly than that of the Brown Corpus, which is reflected in its wider inter-quartile range (transcript 0.26–0.76; BC 0.42–0.55). The observed bridge teams' lexical density differs starkly from that of the SMCP text collection, with PS = 0.30. Had the bridge team only used the SMCP, the effect size would have been 0.50.

Looking at the different content word classes, strong differences between the bridge teams and non-native speakers in a non-nautical setting were found in the distribution of nouns, clear differences in the use of verbs, adjectives and adverbs, and slight differences in numerals. Compared to written texts, bridge teams can be expected to produce far more adverbs and numerals, but fewer nouns, verbs and adjectives.

The SMCP text collection has a strikingly higher proportion of nouns with an effect size of PS = 0.18 and a distinctly higher proportion of verbs. It clearly contains more adverbs and fewer adjectives. The distribution of numerals is nearly identical. In the SMCP, the ratio of nouns is more than twice as high as the corresponding value in the Bridge Team transcript. The unambiguous nature of the SMCP wording is, among others, also reflected in the preference of nouns over determiners such as *this*, *that* or *another*. In the SMCP text collection, only 160 determiners are used which corresponds to a ratio of 0.003, whereas in the Bridge Team transcript, 1,695 determiners were produced, equalling a ratio of 0.039, a value more than ten times higher. Calculating the ratio for nouns and determiners together in both text corpora, the difference is still high (transcript = 0.215, SC = 0.386), but a tendency can be identified to replace nouns with determiners in the transcribed verbal communication. Transcript excerpt 3 offers an example (appropriate SMCP wording is given in brackets).

Transcript excerpt 3

speaker	**utterance**
33) officer	No, no, these, these, these two are cardinal buoys.
	(instead of: We are passing cardinal buoys H and F on port side.)
34) shipmaster	Yes.
35) officer	Then this one is a...
36) shipmaster	Er, tower, whatever.
37) officer	Yeah, but this one is a, is a target.
	(instead of: Sailing boat Alpha is a target.)
38) shipmaster	Yeah, but you showed here, so I was confused.
39) officer	I have such big fingers, so it's very confusing.
40) shipmaster	I think this, this must be the sailing vessel.
	(instead of: The target must be sailing boat Alpha.)
41) officer	Yeah, this one's sailing boat.
	(instead of: Yes, the target is sailing boat Alpha.)

Hypothesis $H_0 4$. The distribution of nautical key words does not differ significantly between bridge team and other, non-nautical communication.

Based on the differences observed in the distribution of content words it was found that utterances by bridge team members present a higher information density than that of speakers transcribed in the Vienna Corpus. However, these findings alone are not suitable to test assumptions related to their appropriateness or maritime idiomaticity. For this reason, a set of key words has been established by isolating all content words included in the SMCP (cf. John et al., 2013). As this method does not include any qualitative differentiation, it does not discriminate between words which are exclusively used in the maritime domain (e.g. anchor, starboard, leeward), those also used outside a maritime setting (e.g. bow, proceed, target) and those without any specific maritime meaning (e.g. dangerous, under, way). Interestingly, this quantitative approach for extracting key words has resulted in a very similar key word ratio (~31%) to the "coverage" reported by Chung and Nation (2003) for anatomy texts. In the following, the empirical data are compared with the two non-nautical text corpora and the SMCP text collection. It is assumed that a higher key word density reflects a higher degree of maritime idiomaticity.

Table 2 shows that the Brown and Vienna Corpus feature a key word ratio of 0.020 and 0.186, respectively. Neither of the two corpora contains specifically maritime texts. The calculated values must therefore be assumed to be the noise generated by the adopted methodology. It is interesting to see that in verbal communication by non-native speakers the proportion of nautical key words is much lower than in texts written by native speakers. This may be explained by the fact that only a small number of the key words are used in normal conversation. To *proceed*, for example, is more likely to appear in written texts than in verbal communication.

The transcribed bridge team communication leads to a key word ratio of 0.307, i.e. roughly one out of three words uttered is a key word. Had the bridge team only used the SMCP, they would have produced a ratio of 0.550 as reflected in the SMCP text collection. Figure 3 displays the key word ratio for the different text corpora as box plots with the inter-quartile ranges in grey (labelled "KW"). The differences in the distribution of the SMCP content words and SMCP key words are caused by the high number of proper names (e.g. for ships, buoys, etc.) which are counted as content words but not as key words.

In order to perform an analysis of the samples' variances, they are tested for a normal and for a homoscedastic distribution. Anderson–Darling's test for normal distribution on each of the four text corpora and Levene's test for homogeneity both resulted in $p < 0.000$, so that neither a normal nor a homoscedastic distribution can be assumed. Again, the non-parametric Kruskal–Wallis test was employed on the four text corpora resulting in $p < 0.000$. Post-hoc Mann–Whitney *U* tests between the Bridge Team transcript and the other three text corpora also lead to $p < 0.000$ for each of them.

For the key word distributions, the following effect sizes were computed:

$$Ps_{KW:Transcript/BC} = 0.59$$
$$Ps_{KW:Transcript/VC} = 0.68$$
$$Ps_{KW:Transcript/SCMP} = 0.13$$

$H_0$4 must be rejected and the alternative hypothesis accepted as the key word distribution between the text corpora differs significantly between the four text corpora. The most striking difference here is not the strong effect sizes between the Bridge Team transcript and the two non-nautical corpora because a much higher proportion of nautical key words is expected in the Bridge Team transcript. It is rather the very strong effect size between the transcript and the SMCP text collection of 0.13. Expressed in real figures this means that in the recorded bridge team communication a high proportion of 2,064 utterances do not contain any key word at all, which equals a ratio of 0.047. The key word median, which covers 50% of all utterances by the bridge team members, equals 0.25, a value which only about 3% of the phrases in the SMCP text collection present. The first three quartiles or 75% of all utterances result in a value of 0.40 which does not even reach the SMCP text collection's first quartile of 0.48. Transcript excerpt 4 and 5 present two speech acts with different key word ratios. Transcript excerpt 4 contains a total of 44 words out of which nine are key words, the key word ratio equalling 0.20. SMCP key words are underlined.

Transcript excerpt 4

speaker	**utterance**
42) shipmaster	What is this? But this is <u>no</u>, <u>no anchorage</u>.
43) officer	Maybe the <u>wrong chart</u>?
44) shipmaster	Yeah, but <u>also</u>.
45) officer	Where is it <u>here</u>? Where is this, where is this <u>anchorage</u>?
46) shipmaster	Yeah, this, this one, or? Neue Weser Reede, maybe? This would be <u>South</u> of Helgoland, but I...

Excerpt 5 consists of 30 words out of which ten are key words leading to a key word ratio of 0.33 which is slightly higher than the observed ratio in the whole transcript.

Transcript excerpt 5

speaker	**utterance**
47) shipmaster	What did you <u>say</u> what was the <u>distance</u> to this one? To <u>make</u> the...
48) officer	Five <u>cables</u>. And <u>then</u> we have to <u>go</u> to the <u>new course</u>.
49) shipmaster	Hm, okay. <u>Steady</u>.
50) officer	<u>Steady</u>.

Hypothesis $H_0$5. The grammar diversity distribution observed in bridge team communication does not differ significantly from other, non-nautical verbal communication.

So far, analyses have been performed on lexical structures including vocabulary growth and word frequencies as well as lexical and key word densities. This last hypothesis deals with possible differences in grammatical diversity between the utterances by bridge team members, the Vienna Corpus and the SMCP text collection. The Bridge Team transcript is also compared to verbatim transcripts of the radio programme Lingua Franca in which a presenter interviews guests, a programme which has been analysed by John and Brooks (2014). For this analysis, the Brown Corpus is disregarded, because here the focus lies exclusively on spoken discourse.

Grammar diversity has been found to vary in dependence on the utterances' length. For this reason, the specific grammar diversity index (spdi) has been developed to determine "the relative deviation of the grammar diversity value in an observed utterance from the expected number given by the PDI [POS Diversity Index] for a specific utterance length" (John and Brooks, 2014, p33).

The grammar diversity displayed as spdi box plots in Figure 4 is more homogeneously distributed than the four lexical distributions analysed above. The Bridge Team transcript's inter-quartile range stretches from 0.83 to 1.05 which is slightly lower than that of the Lingua Franca transcript (0.94–1.12), relatively similar to the Vienna Corpus (0.88–1.02) and slightly higher than that of the SMCP text collection (0.75–1.05). In the Bridge Team transcript and the Vienna Corpus, the median equals 1.00 whereas it is 0.92 in the SMCP text collection and 1.04 in the Lingua Franca radio programme.

A normal distribution was only found for the Lingua Franca transcript. The other two text corpora and the SMCP text collection do not present a normal distribution. Levene's test for homogeneity led to $p < 0.000$, so that no homoscedasticity can be assumed. The Kruskal–Wallis test again resulted in $p < 0.000$. The three Mann–Whitney *U* tests performed post-hoc between the Bridge Team transcript and the other text corpora also resulted in $p < 0.000$; the null hypothesis $H_0 5$ is therefore rejected.

For the spdi distributions, the following effect sizes were computed:

$Ps_{spdi:Transcript/LF} = 0.35$
$Ps_{spdi:Transcript/VC} = 0.65$
$Ps_{spdi:Transcript/SCMP} = 0.56$

The grammar diversity in bridge team communication by non-native speakers resembles the speech acts by non-native speakers outside a maritime setting as transcribed in the Vienna Corpus. Although the statistical testing has resulted in a rejection of the null hypothesis, the small effect size between the two non-native speaker's corpora of $PS = 0.65$ nevertheless indicates a close proximity of the observed grammar structures. The grammar structures in the SMCP text collection presents an even higher degree of similarity with $PS = 0.56$. To the contrary, the transcripts of the Lingua Franca radio programme differ more strongly with $PS = 0.35$ pointing at more elaborate speech acts.

4. Limitations of the adopted methodology

No research design is totally free of a certain degree of limitations and biases. This research studies transcribed verbal communication while disregarding any para-verbal (speed, intonation, etc.) and non-verbal communication (eye movement, facial expression, gestures, etc.). It is based on simulation sessions carried out with volunteering university students of Nautical Sciences. Here, a general symmetry in speech facilitates an ideal speech situation in which bridge team members participate under an assumption of equality (Habermas, 1979) which will not always be the case on board a sea-going ship. Another difference occurs due to the fact that only one team was inter-cultural while the other bridge teams were all mono-cultural. The introduced sampling bias had to be conceded by the authors because a data sampling in real-life conditions is virtually impossible.

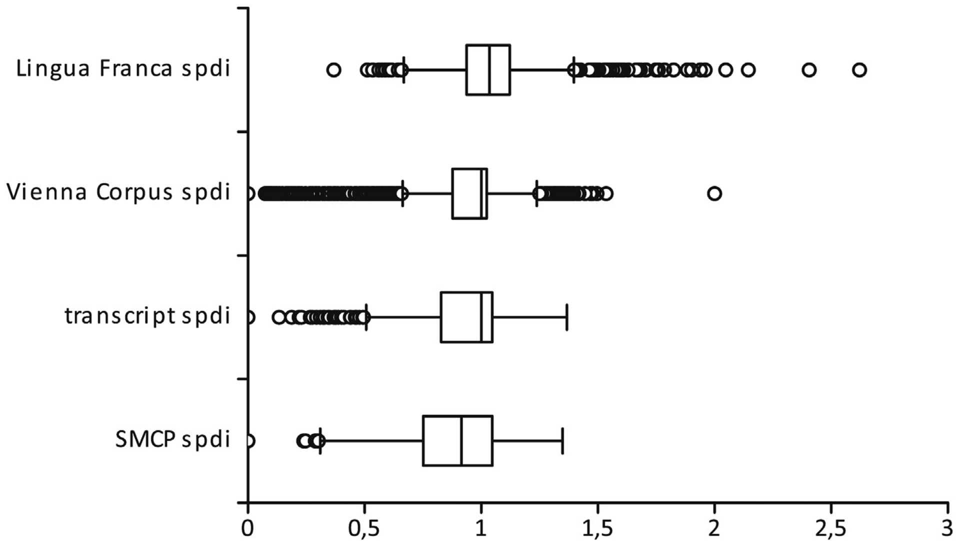

Figure 4. Special POS diversity index for analysed text corpora.

With regards to the samples' representativeness, the bridge team communication sampled from the German teams can only be considered representative of a German group of prospective nautical officers with a prolonged English language learning experience. However, the authors would like to argue that this group can also be considered representative for all northern European non-native speakers of English with a similar language exposure in school and university education.

5. Findings and discussion

This research identifies lexical differences and similarities of quantitative linguistic features between the specific ESP setting of bridge team communication and other text corpora in order to profile a non-native bridge team's idiosyncratic speech patterns. For this purpose, five null hypotheses were tested for significant differences, and inferences were made to quantify the observed statistics.

It was found that the inter-textual vocabulary growth ($H_0 1$) of bridge team communication is not only significantly gentler than in written English but also more gradual than in spoken English outside a maritime setting. In a transcript consisting of 43,019 tokens, only 1,843 types can be expected, i.e. a type-token ratio (TTR) of 0.042 (Brown Corpus = 0.179, Vienna Corpus = 0.071). For any given number of tokens (n) up to 43,019, the corresponding number of types (V) can be computed by using the power function $V(n)_{BT} = 0.02n^{1.91}$ ($R^2 = 0.995$). A comparison with the SMCP text collection validated the similarities of the natural speech patterns with the mandatory, coded SMCP language with its 1,883 word types and a type-token ratio of 0.040.

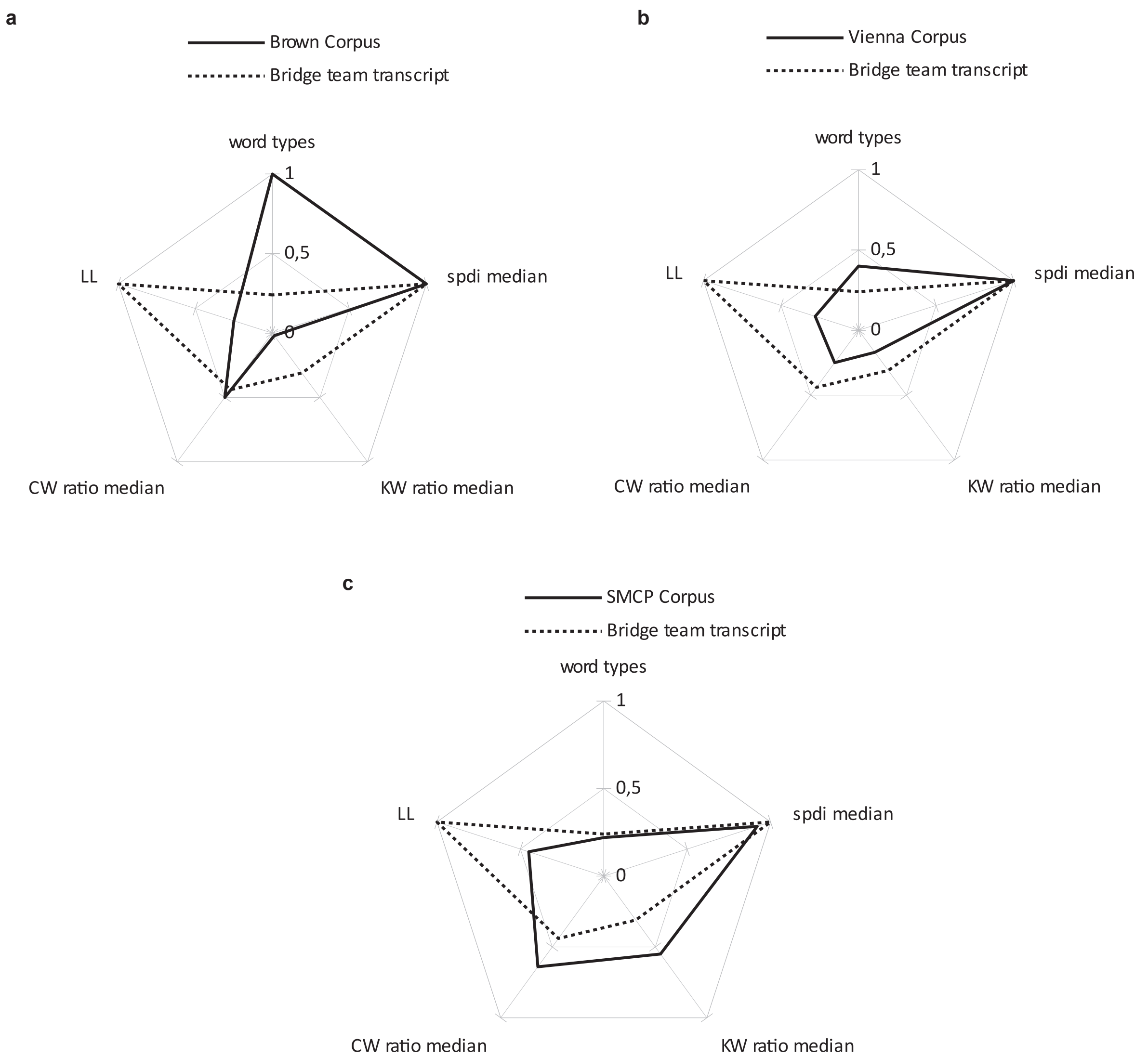

Figure 5. Linguistic profile comparison of the analysed text corpora.

Relative word frequencies (H_02) were also found to differ significantly between bridge team communication and the other analysed text corpora. The four corpora only shared 406 word types of which the Brown Corpus included 305 with a significantly different distribution. Again, this Corpus presented the highest difference with a median log likelihood statistic of 18.92 and only 21% of all shared word types presenting a similar frequency. For the Vienna Corpus, the corresponding values were 15.09 and 28%. The bridge team transcript is much closer again to the SMCP text collection with a median log likelihood statistic of 8.73 and 44% of word types occurring at a similar relative frequency.

A significantly differing distribution of content words (H_03) was ascertained between the Bridge Team transcript and the other corpora. In total, the content word ratio of the speech acts by the bridge team members was more similar to the written English reflected in the Brown Corpus, with an effect size of PS = 0.49 as opposed to PS = 0.70 for the Vienna Corpus and PS = 0.30 for the SMCP text collection.

Nautical key words (H_04) are assumed to indicate the appropriateness or idiomaticity of bridge team communication. The transcript's key word ratio of 0.307 is smaller than that of the SMCP text collection (0.550) but higher than in the Brown and Vienna Corpus (0.020 and 0.186, respectively). Nevertheless, the bridge teams' key word distribution differed significantly from the other corpora and surprisingly the strongest effect size was observed in the comparison with the SMCP text collection, with PS = 0.13.

To compare the Bridge Team transcript's special POS diversity index (H_05), the Brown Corpus was replaced by the transcript of the Lingua Franca radio programme analysed by John and Brooks (2014). A significantly different distribution was found between the corpora although the effect sizes between the Bridge Team Transcript and the SMCP text collection were relatively small in comparison to the Vienna Corpus and Lingua Franca programme.

So far, the dependent linguistic variables have been analysed individually. However, by combining them, a *linguistic profile* is created which effectively quantifies the observed language patterns as a coherent whole. This *profile* is presented graphically in Figure 5 as a five-pointed polygon in which the variables have been normalised for a clear visualisation. The overlapping areas underline the much closer proximity between the Bridge Team transcript and the SMCP text collection than to the other two text corpora. The figure also shows that the observed bridge team communication still differs clearly from the mandatory coded language of the Standard Marine Communication Phrases.

Profiling the inherent communication patterns of the collected speech samples leads to a quantitative model for the specific discourse community of bridge team members in full-mission simulation exercises speaking English as their second language. This research has shown that the created model is a valid tool for quantifying the differences and similarities between the sub-genre of bridge team communication and other communicative settings.

Future research should profile the idiosyncratic language patterns of differently composed bridge teams (e.g., native English speakers) for comparison with the profile represented in this paper. Such a comparison will identify differences in the discourse between non-native and native English speakers in this particular environment, and those results may be generalisable to work teams in other domains. Correlating the profile with other behavioural data that identifies the quality of the communication may also be useful. Finally, this may lead to a model of standard communication behaviour which can be used as a benchmark in the training of future nautical officers. Communication is a decisive component of social interaction and a contributory factor to improve safety at work. In shipping, improved education and training in bridge team communication will take us a step closer towards avoiding fatal accidents as in the case of passenger vessels Costa Concordia, Sewol and many others.

Funding

The corresponding author has received a PhD scholarship from the Australian Maritime College, an institute of the University of Tasmania, Australia.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

The authors thank Prof Hans-Jörg Nafzger, Klaus Damm, Christian Sievers and Thorsten Kramer of the Maritime Faculty in Elsfleth for their kind support in the collection of team interaction speech data.

References

Baker, P., & Ellece, S. (2011). *Key terms in discourse analysis*. London: Continuum.
Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language, 15*(2), 103-116.
Cole, C., & Trenkner, P. (2012). Whither maritime English? – 2012. In *Proceedings of the international maritime English conference (IMEC)* (pp. 3-18). Yangong: Maritime Lecturers' Association.
Covington, M., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type-token ratio MATTR. *Journal of Quantitative Linguistics, 172*, 94-100.
Deboo, K. M. (2004). Maritime resource management online. Alert. Internet resource available at http://www.he-alert.org (Viewed on 18 April 2014).

Dževerdanović-Pejović, M. (2013). Discourse of VHF communication at sea and the intercultural aspect. *International Journal for Traffic and Transport Engineering, 34*, 377-396.
Francis, W. N., & Kucera, H. (1979). Brown corpus: A standard corpus of present-day edited American English, for use with digital computers. In *Department of linguistics, Brown University. Revised 1971, revised and amplified 1979*. Providence: Rhode Island.
Gregorič, T., & John, P. (2013). SMCPexamples – learn SMCP phrases by examples. Internet resource available at http://www.smcpexamples.com (Viewed on 30 June 2014).
Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research. Univariate and Multivariate Applications*. Hove: Taylor & Francis.
Habermas, J. (1979). *Communication and the evolution of society* (Vol. 29)Boston: Beacon Press.
Halliday, M. A. K. (1987). Spoken and written modes of meaning. In R. Horowitz, & S. J. Samuels (Eds.), *Comprehending oral and written language* (pp. 55-82). San Diego: Academic.
Horck, J. (2005). Getting the best from multi-cultural manning. *BIMCO Bulletin, 100*(4), 28-36. London: Baltic and International Maritime Council.
Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*(2), 236-248.
International Maritime Organization. (1995). *Standards of training, watchkeeping for seafarers, as amended*. London: International Maritime Organization.
International Maritime Organization. (2001). *Resolution A.91822: IMO standard marine communication phrases*. London: International Maritime Organization.
Jensen, O. C., Sørensen, J. F., Thomas, M., Canals, M. L., Nikolic, N., & Hu, Y. (2006). Working conditions in international seafaring. *Occupational Medicine, 566*, 393-397.
John, P., & Brooks, B. (2014). Lingua franca and its grammar footprint: Introducing an index for quantifying grammatical diversity in written and spoken language. *Journal of Quantitative Linguistics, 211*, 22-35.
John, P., Brooks, B., Wand, C., & Schriever, U. (2013). Information density in bridge team communication and miscommunication – a quantitative approach to evaluate maritime communication. *WMU Journal of Maritime Affairs, 122*, 229-244.
Kilgarriff, A. (2005). Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory, 12*, 263-276.
Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics, 192*, 255-271.
McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.
Möckel, S., Brenker, M., & Strohschneider, S. (2014). Enhancing safety through generic Competencies. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation, 81*, 97-102.
Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. *Handbook of Reading Research, 3*, 269-284.
Noble, A., Vangehuchten, L., & Van Parys, W. (2011). Communication for maritime purposes: Some exploratory results of a survey-based study on intercultural and linguistic features. *International Journal of Applied Linguistics, 162*, 111-133.
Potter, R. E., & Balthazard, P. A. (2002). Understanding human interaction and performance in the virtual team. *Journal of Information Technology Theory and Application JITTA, 41*, 6-30.
Pritchard, B. (2003). Maritime English syllabus for the modern seafarer: Safety-related or comprehensive courses? *WMU Journal of Maritime Affairs, 2*(2), 149-166.
Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In *Proceedings of the workshop on comparing corpora* (pp. 1-6). Stroudsburg: Association for Computational Linguistics.
Seidlhofer, B., Breiteneder, A., Klimpfinger, T., Majewski, S., Osimk-Teasdale, R., Pitzl, M. L., & Radeka, M. (2013). The Vienna-Oxford International Corpus of English version 2.0 XML. Internet resource available at http://www.univie.ac.at/voice/ (Viewed on 30 June 2014).
Trenkner, P. (1996). IMO-Standard Marine Communication Phrases (SMCP) – an attempt to meet increased communication requirements of ship's officers. In *Ninth international maritime lecturers' association international conference on maritime education and training (IMLA 9)*. Kobe: International Maritime Lecturers' Association.
Trenkner, P., & Cole, C. (2010). Raising the maritime English bar: The STCW manila amendments and their impact on maritime English. In *Proceedings of the 22nd international maritime English conference* (pp. 3-16). Alexandria: International Maritime Lecturers' Association.
Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading, 151*, 8-25.
Weeks, F. (1997). Whither maritime English. In *Ninth international IMLA workshop on maritime English (WOME 9)*. Malmö: International Maritime Lecturers' Association.

Peter John is a senior lecturer of English and Spanish. He has been teaching Maritime English since 1996 and is currently undertaking his doctoral research on maritime communication. He is a member of the Paper and Activities Committee of the International Maritime English Conference (IMLA-IMEC).

Dr Benjamin Brooks is a human factors psychologist and Senior Research Fellow at AMC, University of Tasmania. His research examines human performance in high risk work environments and operational and educational technologies to maintain & improve this performance.

Dr Capt Ulf Schriever is a senior lecturer of Navigation and Maritime Communication at AMC, University of Tasmania. He combines his teaching with periods of working on board sea-going ships as navigational officer or shipmaster.