English for Specific Purposes 66 (2022) 111–130

Contents lists available at ScienceDirect

English for Specific Purposes

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

Technical single and multiword unit vocabulary in spoken rugby discourse

Stuart Benson [a,*], Averil Coxhead [b]

[a] Center for Language Research, University of Aizu, Japan
[b] School of Linguistics and Applied Language studies, Victoria University of Wellington, New Zealand

ARTICLE INFO

Article history:
Available online 18 February 2022

Keywords:
Rugby union
Sport
Technical vocabulary
Multiword units
Vocabulary knowledge

ABSTRACT

Rugby union is played by first and second language speakers of English in many countries. Rugby has its own technical vocabulary (e.g., *ruck*, *inside pass*) and its knowledge is crucial for joining the rugby community. This article reports on a study into the nature and knowledge of technical vocabulary in spoken rugby discourse through identifying technical vocabulary and developing both single and multiword word lists. This study involved the creation of a spoken rugby corpus (61,295 running words) and a vocabulary load analysis using Nation's (2012) BNC/COCA frequency and supplementary lists. The results found learners need over 4,000 word families plus supplementary lists to reach 98% comprehension. The corpus analysis resulted in a list of 252 technical single words which covers over 12% of the spoken corpus. A pedagogically oriented multi-word unit list of 267 items was also created. A receptive knowledge task using a sample of the rugby lexis and general English words was administered to 77 (29 English L1; 48 L2 speakers of English) participants. The results highlighted differing levels of technical rugby vocabulary knowledge between first and second language speakers of English. This article concludes with implications for ESP rugby courses and suggestions for possible future research.

1. Introduction

Rugby union (hereafter referred to as 'rugby') is played by people in many countries where the first language (L1) is English. Since becoming a professional sport in 1995, rugby has gained popularity in English as a second or foreign language countries (L2), in countries such as Japan (Chiba & Jackson, 2006; Harris, 2010). Spoken communication with players, coaches, managers, and the referee is essential in the sport. For second language learners, this predominately spoken environment presents a challenge for comprehension and communication in a team-based and, at times, highly charged environment. Previous research notes the prominence of technical vocabulary, such as *ruck, maul,* and *lineout*, in spoken rugby discourse (Kuiper & Lewis, 2013; Wilson, 2011); however, there is little research in vocabulary in rugby beyond this.

This article investigates specialised vocabulary in spoken rugby discourse from several perspectives. Firstly, it presents a frequency-based analysis of spoken rugby corpora to find out how many word families learners need to know in English to cope with speaking and listening in the sport. This is important because L2 speakers who move countries to play rugby need to have a baseline of vocabulary knowledge to understand what players and coaches are saying in English in training, during

* Corresponding author. C/. The University of Aizu, Tsuruga, Ikki-machi, Aizuwakamatsu, Fukushima 965-8580 Japan.
E-mail addresses: benson@u-aizu.ac.jp (S. Benson), averil.coxhead@vuw.ac.nz (A. Coxhead).

https://doi.org/10.1016/j.esp.2022.02.001
0889-4906/

games, and at practice. Secondly, the article outlines the development of principled technical word lists in spoken rugby based on types (not word families) and the proportion of this vocabulary in the corpus. This research is important because it will help L2 players and their coaches identify which words are worth learning for rugby. This research is methodologically innovative because of its numerous approaches to the data to identify this lexis. Thirdly, it provides a multiword unit list, based on the spoken rugby word list, which include items such as *front row*, *short pass* and *behind the scrum*. The last piece in the puzzle is a small-scale study of the knowledge of rugby specific vocabulary of English L1 and Japanese L1 players and coaches. This work highlights the need for learning rugby-specific terms in English. Together, these perspectives provide valuable insights into the vocabulary of rugby to assist learners and English for Specific Purposes (ESP) teachers to address a major gap in sports-related studies in English.

1.1. Vocabulary in sports language: football and rugby

The most researched sport in the literature is football (Lavric, Pisek, Skinner, & Stadler, 2008). Such studies have primarily focused on word strings or multiword units, such as *corner ball* or *to score a goal*. They have also examined the vocabulary used in sub-domains, such as TV, radio, and online commentary (Bergh, 2011; Humpolík, 2014), and written match reports (Schmidt, 2008). Bergh and Ohlander (2012, 2017) selected 25 English football words (e.g., *match, corner,* and *dribble*) that were "central to the football domain" (Bergh & Ohlander, 2012, p. 287), from *A Dictionary of European Anglicisms* (Görlach, 2001) and investigated their influence in 16 European languages. This research is valuable because it highlights the kinds of technical words in particular sports, and the spread of this vocabulary in the 'beautiful game'. Rugby has also been the subject of such research, but to a lesser extent, and its main focus has been the discourse of team-based speech (Wilson, 2009a; 2009b, 2011) and TV rugby commentary (Desmarais & Bruce, 2009, 2010; Kuiper, 1996; Kuiper & Lewis, 2013). While vocabulary has been noted as a prominent feature of the sport (Kuiper & Lewis, 2013; Wilson, 2011), these previous studies have not addressed the needs of L2 players and their coaches. The main aim of the present study is to identify which words are worth learning for rugby speakers.

1.2. Vocabulary knowledge and spoken language

How many words do L2 learners need to know to cope with the demands of spoken English?

Vocabulary load studies recommend 95% and 98% as minimal and optimal comprehension thresholds for spoken texts respectively. van Zeeland and Schmitt (2013) examined listening comprehension of short stories and found 95% coverage provided "good but not necessarily complete" comprehension, whereas 98% provided "very good comprehension" (pp. 18–19). Much of vocabulary load research to date has focused on everyday English, academic and ESP contexts. Findings include that general spoken English requires 2,000–3,000 word families for 95% and 6,000–7,000 word families for 98% coverage (Nation, 2006). Dang and Webb (2014) found that academic spoken English requires 3,000–5,000 word families (95%) and 5,000–13,000 word families (98%), but there is some disciplinary variation with social sciences having the lowest vocabulary load (5,000 word families at 98%) compared to life and medical sciences (13,000 word families at 98%). In trades education-based research, Coxhead and Demecheleer (2018) found that 95% coverage over a corpus of tutor talk in plumbing trades education was reached by 3,000 word families plus proper nouns and marginal words, and 98% coverage was reached by 5,000 word families plus proper nouns and marginal words.

1.3. Technical vocabulary

Developing knowledge of technical vocabulary in a specialised field is vital for any learner, yet little is known about the size of this lexis and how it grows over time (Coxhead, 2018; Nation, 2016). Players and coaches who have been involved in rugby (or interested bystanders and fans) over a long period of time develop a strong understanding of the technical vocabulary of the game, in English in their L1 or L2. If there is a difference in the knowledge of technical vocabulary between English L1 and L2 rugby players, there is a stronger basis for suggesting that this vocabulary is needed to prepare these players if they move to an English L1 environment, or perhaps even to be able to understand and participate in talk about rugby in English – something that 'rugby heads' in any language absolutely love to do. The primary audience for this research is L2 speakers of English who need to learn rugby lexis to play or coach in English, and also, potentially, L1 speakers who might be coaching such players.

Technical vocabulary relates closely to a subject area (Nation, 2013; Woodward-Kron, 2008). It is difficult to talk about a sport such as rugby without mentioning moves (e.g., *tries* and *lineouts*) or player positions (e.g., *lock*, *winger*) that are central to the game. Knowing and using the technical vocabulary of a domain is an important indicator of belonging to a group who uses that same language (Wray, 2002) as can be seen in research in trades education in areas such as carpentry (Parkinson & Mackay, 2016). This means that knowledge of technical rugby vocabulary can assist non-native speaking players and coaches in learning about the game, participating in the sport, and becoming part of a community.

A common misconception about technical vocabulary is that it is made up of words that are not very common. However, these words can be high, mid, or low frequency (Nation, 2016). This point is important, because everyday words may also be technical, e.g., *flow* in plumbing (Coxhead & Demecheleer, 2018) and early studies in EAP such as Coxhead (2000) did not

include high frequency words even though they are potentially high frequency technical items in academic English (see Dang, Coxhead & Webb, 2017). This means that in rugby, some technical items may occur in everyday English and in rugby can occur with the same meaning (e.g., *ball*, *angle* and *low*) or with different meanings (e.g., *knock on*, *match*). A second key point to make here is that some lexical items may be highly technical, that is, occurring only in rugby (e.g., *loosies* and *ruck*). For example, Ha and Hyland (2017) found a relatively small number of highly technical words in a study of spoken finance lexis. These items need to be taken into account in any study and are possibly the most easily recognised as being technical because of their narrow range of use.

Technical vocabulary can make up a substantial amount of words in technical written texts. Chung and Nation (2003) identified the technical vocabulary in an anatomy textbook (Chung is a medical professional and applied linguist) using a semantic scale and found 31.2% of the total tokens (running words) were technical. This amounts to roughly one word in three. Similar levels have been found in medical textbooks (37%) (Quero, 2015), pedagogical plumbing texts (35.58%) (Coxhead & Demecheleer, 2018), and pedagogical fabrication texts (30.47%) (Coxhead, McLaughlin & Reid, 2018). Depending on the discipline, coverage of technical vocabulary in spoken discourse is approximately 10%, which means it is roughly one third of the technical vocabulary in written texts (Coxhead et al., 2020a). Rugby players tend to learn through communication rather than reading, so it is important to establish the size of that learning goal. With the large amount of research highlighting the importance of technical vocabulary in a specialised area, it is necessary to investigate 1) the nature of technical vocabulary in spoken rugby discourse, and 2) the knowledge of this technical vocabulary. By understanding L1 and L2 rugby players' knowledge of technical rugby vocabulary, appropriate measures can be put into place within an ESP rugby language class to remedy the issue.

1.4. Word lists of technical vocabulary

Word lists are useful tools for presenting technical vocabulary for language learning and teaching. Discipline-specific word lists can help with setting learning goals, developing curricula, and testing (Nation, 2016). Word lists based on spoken corpora are less common than those on written corpora, and most spoken word lists have focused on the university context (Dang, 2018a, 2018b; Dang et al., 2017). Most technical word lists use *types* (individual words) as the unit of counting words used rather than word families (Chung & Nation, 2003, 2004; Coxhead et al., 2016; Coxhead & Demecheleer, 2018, 2018; Quero, 2015; Watson-Todd, 2017). This is because individual types may be technical but members of a word family may not have a technical meaning. For example, *advantage* (with a similar meaning to football or hockey) is technical in rugby, but *advantageous* is not.

Technical items in word strings or multiword units (MWU) are important for learners (see Coxhead, 2019). Word lists of these items can draw the attention of learners and teachers to "… productive patterns which are tied to specific lexis in a way that can lead them to be overlooked by traditional grammars" (Durrant, 2009, p. 163). Some examples of technical multiword units in ESP include *black box, base leg, sniffer dog* in aviation (Aiguo, 2007) and *present with* in medical communication (Basturkmen, 2010). Multiword unit word lists have been developed in engineering (see Fox & Tigchelaar, 2015; Wood & Appel, 2014 for example). A technical word list would be useful for L2 learners of English, for example in countries such as Japan, because they may play rugby with or be coached by L1 speakers of English, or choose to play rugby in an English-speaking country, such as New Zealand.

In this article, five main categories of rugby-specific vocabulary are explored: (1) rugby words used with the same meaning as everyday usage; (2) rugby words used with a meaning different to that of everyday usage; (3) rugby-specific words; (4) standalone MWUs; and (5) variable MWUs.[1]

2. Background and context

Firstly, rugby involves 15 players on each side, with one referee on the field and two on the sidelines. Games are 80 min long but may continue beyond that time limit depending on the game. Secondly, the spoken corpora in the present study were gathered in the New Zealand rugby context and the receptive knowledge task took place in New Zealand and Japan. These two countries were chosen because there is a rich association in coaching and playing rugby between them and the first author has played rugby in both countries. The project investigated three key issues: 1) the lexicon of spoken rugby discourse, 2) the technical vocabulary in spoken rugby discourse, and 3) the linguistic needs of foreign rugby players.

The following research questions guide this investigation:

1. What is the vocabulary load of spoken rugby discourse?
2. What is the coverage of the single word technical spoken rugby word list over the spoken rugby corpus?
3. What are the technical multiword units in the spoken corpus?
4. To what extent might the receptive knowledge of technical rugby vocabulary differ between L1 and L2 rugby speakers?

[1] Thank you to the anonymous reviewers for these suggestions.

3. Methodology

3.1. Spoken rugby corpora

The goal of this study was to investigate the vocabulary in spoken rugby discourse as it is the primary type of communication in the sport. With over 25 years' experience playing and coaching the game, the lead author understands rugby to be an interactional sport, in which spoken discourse and effective communication between players and coaches are essential to playing. Furthermore, rugby is not only a performance which is carried out, but a sport which is commonly observed, whether attending a game or watching on television (TV). As commentary can enhance the visual experience (Desmarais & Bruce, 2010), it is used to bring meaning to what is occurring on the television screen.

A spoken corpus containing authentic spoken discourse from two sub-domains in rugby was created. Table 1 provides an overview of the corpus. The first sub-domain drew on Wilson's (2011) corpus of authentic interactions in a rugby setting which contained 25,637 running words. The interactions were recorded in 12 situations: interviews, training, post-match team room, pre-match huddle (the players and coaches gather together in a circle), half-time huddle, full-time huddle, team meeting, warmup, front-row (players who make up the first row in a scrum) warmup, pre-warmup locker room, water message (the person who brings out water to the players delivers instruction from the coach), and full match day. The transcribed interactions totaled 2 h and 10 min.

The second corpus was a TV commentary corpus containing 35,658 running words from three orthographically transcribed rugby games. The three games were played in the New Zealand Super Rugby competition and the commentators were well-known New Zealand television rugby commentators: Grant Nisbett and Justin Marshall. Only commentary during the 80 min of the games was transcribed, resulting in 240 min of commentary. Automatic transcription software was trialed to ascertain if it was a viable means of increasing the size of the corpora. However, the software was only able to successfully transcribe 25% of the commentary. Speed of the commentary, background noise, accent of the commentators, and Pasifika names were all possible causes for errors. Therefore, manual transcription was the best option for creating the corpus. It is important to sound a note of caution with this commentary corpus in the two commentators may have their own preferred ways of expressing themselves.[1] These commentators have different roles in commentary: one covers the game and the other provides 'colour' including expert analysis of the game, statistics and background information. Justin Marshall is a former All Black who provides colour, and Grant Nisbett is a professional broadcaster.

The decision to create the TV commentary corpus to represent spoken rugby discourse was based on two reasons. First, adding to Wilson's (2011) corpus was problematic because of the inaccessibility of collecting authentic interactional rugby speech. In addition, such speech contains a large amount of sensitive data, such as codes for particular moves, which cannot be included to preserve the integrity of each team (Wilson, 2011). To protect the data, this speech needs to either be deleted or scrambled, which in turn changes the representativeness of actual language in use. Second, with rugby being broadcast in foreign speaking countries, such as Japan, TV commentary is one of the most accessible methods for second language learners to hear spoken rugby discourse. Furthermore, as this commentary language feeds back into how coaches and players talk (Wilson, 2011), namely positions of players and set pieces within the game, the pedagogical implications of using TV commentary was an influential factor when building the corpus.

The spoken rugby corpus is relatively small compared to other specialized spoken corpora (see Coxhead, Demecheleer & McLaughlin 2016), and only contains discourse from two sub-domains. It is also important to note that these corpora are not the same size, but they both represent interactional talk about rugby. The corpora were joined for our analysis. Clearly a larger corpus would provide a bigger picture of rugby language. A written corpus, for example, could have potentially yielded a larger pool of technical vocabulary, but rugby players and coaches do not draw heavily on written documents. The aim here is to investigate vocabulary in use in spoken rugby and to use the corpus to develop discipline-specific word lists. We also hope to demonstrate that this research can help shed some light on the technical lexicon in under-researched domains.

Table 1
Overview of the spoken rugby corpus.

Corpora	Time (minutes)	Total running words
Interactional corpus	130	25,637
TV commentary corpus	240	35,658
Total	370	61,295

Cleaning the corpus was necessary in order to conduct the analysis. Contractions (e.g. *'cause, didn't*) were expanded (e.g., *because, did not*). Unfinished words, such as 'wa-' (from 'way') and 'ver-' (from 'very'), were excluded even though they may assist the listener's comprehension in spoken contexts (Harris, 2003). Following Coxhead and Demecheler (2018), once the analysis was complete, contractions were returned to their original state to keep the corpus representative of the language in use.

3.2. Corpus analysis

The analysis involved two procedures. First, all the words in the spoken corpus were categorised. Second, several principles were followed to develop the single word technical spoken rugby word list.

3.2.1. Vocabulary load of spoken rugby discourse

The *Range* program (Heatley, Nation & Coxhead 2002) was used to conduct the vocabulary load analysis. This program allows users to enter written texts and analyse its lexical coverage according to Nation's (2012) frequency-based 25,000 word lists from the British National Corpus and the Corpus of Contemporary American English (BNC/COCA). The program also presents the frequency and coverage of each word according to the word lists. Nation's (2012) BNC/COCA word lists and four supplementary word lists of proper nouns, compounds, abbreviations, and marginal words were adapted for the analysis.

This procedure entailed:

1. Adding word family members to their existing word families in the BNC/COCA lists, if they satisfied Bauer and Nation's (1993) word family scale. For example, *kickable* was added to the word family *kick*. In total, 21 types were added to their corresponding word families.
2. Adding proper nouns not found in the existing supplementary list. If the meaning of the noun only occurred in the corpus as a proper noun, it was deleted from its base list and subsequently added to the supplementary list: E.g., *Crusaders, Zoo, Panthers.* If the meaning of the noun occurred as both a proper and common noun, it was not added: E.g., *Weeks, Brown, Force.* In total, 198 proper nouns were added to the supplementary proper noun list.
3. Adding abbreviations to the supplementary list. In total only four occurred in the spoken corpus, e.g., OBU: Old Boys University, MSP: Marist St. Pats, TMO: Television Match Official, and NRL: National Rugby League.
4. Creating an additional base word list for words that did not occur in Nation's (2012) BNC/COCA and supplementary word lists. These items were kept in their own base list because they were likely to be technical rugby words and therefore candidates for the rugby word lists. In total, there were nine items (see Table 5 for examples). This list was named *additional base word list.*

Note that this procedure did not involve the BNC/COCA corpora themselves, but the frequency and supplementary lists developed by Nation from those corpora.

3.3. Identifying technical rugby vocabulary in the corpus for the word list

A mixed method quantitative and qualitative approach was used to identify technical single words and multiword units in the corpora and subsequently create the technical spoken rugby word list. First, two frequency principles were applied to the results of the *Range* program (Heatley et al., 2002). A preliminary analysis of multiple cut-off points from 10 to four were initially applied to identify possible technical words in the spoken rugby corpus. Through this provisional analysis, the final frequency principles applied to the analysis were:

1. Words occurring more than seven times in the rugby corpus from Nation's (2012) BNC/COCA base and supplementary word lists.
2. Words occurring more than four times in the additional base word list.

The frequency principles resulted in cut-off points which allowed the maximum number of potential technical words to be identified. Once clearly non-technical words, such as function words (e.g., *the, and*) were removed, the frequency analysis identified 363 items that were potentially technical rugby words.

Second, a qualitative analysis of using a semantic rating scale created by Chung and Nation (2003) and adapted by Quero (2015) was further adapted for this analysis to distinguish whether a word is technical, and to what degree it is specialised within the rugby context (see Table 2). As can be seen in Table 2, non-technical items are in scale 1 while scales 2, 3, and 4 are for technical words.

Table 2
Adapted semantic rating scale with examples.

Scales	Description	Scale	Examples
Non-technical	A word that is general and not related to rugby	1	*like, here*
Technical	A word that is used in rugby but with the same meaning most frequently encountered in everyday usage	2	*attack, meters*
	A word that is used in rugby, but with a particular meaning not frequently encountered in everyday usage	3	*season, wing*
	A word that is unique to rugby and only associated with rugby	4	*loosehead, ruck*

The semantic rating scale descriptors were adapted because they needed to be used by three experienced rugby players when they rated the technicality of the potential items for the lists. Simplifying the descriptors can also lead to increased inter-rater reliability (Chung & Nation, 2004). The three raters had over 15 years of experience in rugby and were trained in using the scale before rating the 363 items from the frequency analysis. The raters were asked to use their knowledge of rugby to rate each item and not refer to dictionaries or other materials. Once ratings were complete, the results were collected and compared. A total of 40 items from scale 1 were not rated as technical, resulting in a reduced primary list of 313 technical spoken rugby words. This list was then analysed for related types; that is, words that are both technical and related (Coxhead,

2000) such as *meter* and *meters.* From a pedagogical standpoint, keeping such items separate in the final word list would create an unnecessary learning burden or confusion if these words were listed separately in a word list. Therefore, while consulting the corpus, the lists were classified for related word types. The two principles followed were:

1. The words are semantically the same.
2. They received the same semantic rating in terms of technicality.

When related word types were found, the word frequencies were combined. This meant that the base form of the word was noted and the word part of the related type was added (e.g., *meter – s*). Once this classification was complete, the spoken rugby word list contained a total of 252 items. It was then organized into groups according to technicality (see Appendix A).

3.4. Developing the spoken rugby multiword unit list

Once the single spoken rugby word list was created, four principles were followed to identify technical MWUs in the corpus:

1. Length: Two to five word units.
2. Unit of counting: Word type.
3. Frequency: Five or more times in the corpus.
4. Technicality: Contains a technical word from the single spoken rugby word list.

As with the single rugby word list, a preliminary analysis of multiple frequency cut-off points was carried out to identify possible MWUs. Based on this analysis taking into account the small size of the corpus, a frequency threshold was set to five times in the corpus. A provisional list of 414 MWUs was created. Further analysis showed that there was considerable overlap between two, three, or four word units. For example, to ki*ck* (2) appears in *the intention to kick* (4), and *intention to kick at goal* (5). To solve this problem and create a more pedagogically applicable MWU list, Wood and Appel's (2014) approach of highlighting the root structure of the MWU and placing any variable slots in brackets at either end of the structure was followed. The frequency of each MWU was also checked for the root structure and the complete MWU. For example, '*to ground*' occurred 37 times in the provisional MWU list. Therefore, in MWUs such as, '*brought to ground*', '*first to ground*', and '*first to ground the ball*', the root structure was '*to ground*' and the other words (e.g., *brought, first, the ball*) were identified as variable slots. By identifying the root structure and the complete MWU, the number of MWUs in the final list reduced considerably. In total, the spoken MWU list contains 267 MWUs (see Appendix B).

3.5. Receptive knowledge task

A receptive knowledge task of 30 lexical items (15 technical rugby words and 15 non-rugby specific words as distractors) was developed. Technical items were chosen according to their semantic rating scale rating, with five items from each of the three scales. The 15 general English items were from the high frequency first 3,000 words in Nation's (2012) BNC/COCA 25,000 word lists, so participants were more likely to recognise the words (e.g., *sick*). None of the general English items occurred in the rugby corpus. Items in the receptive knowledge task are shown in Table 3.

Data gathering was conducted using an online survey as part of a needs analysis investigating the linguistic needs of players and coaches in New Zealand and Japan. Participants were recruited using a snowball technique (Browne, 2005) and a total of 77 responses from 29 English L1 speakers and 48 Japanese L1 speakers who had coached or played rugby for at least one season responded and consented to take part in the study. A Mann–Whitney U test was conducted to compare the sample means, which negated the disparity between of the two groups. The 77 participants were asked to highlight words from the list which they believed to be closely related to rugby. The task was scored simply by adding up the items that each participant indicated was known.

Table 3
Items from the receptive knowledge task.

Technical rugby words	General English words
back	teach
defence	concentrate
penalty	fashion
momentum	journal
intercept	border
jersey	sister
winger	tent
scrum	agriculture
offload	sick
ruck	sew
halfback	negotiate
lineout	neighbor
tighties	garage
breakdown	virus
loosies	letter

4. Results and discussion

4.1. Research question one: What is the vocabulary load of spoken rugby discourse?

Table 4 shows the coverage of Nation's (2012) BNC/COCA frequency and supplementary lists and the additional base word list over the spoken rugby corpus. The coverage reaches the minimum comprehension threshold of 95% at 3,000 word families plus the four supplementary lists and the more ideal threshold of 98% at 4,000 word families plus the supplementary lists. These results are in line with other studies that investigated the vocabulary load in spoken contexts (Coxhead & Demecheleer, 2018; Webb & Rodgers, 2009a; 2009b).

Table 4 shows the coverage of the corpus across Nation's BNC/COCA frequency and supplementary lists and the additional base word list which captured all rugby-specific vocabulary which did not occur in Nation's BNC/COCA lists. The table shows that the highest coverage (84.34%) in the rugby spoken corpus is provided by the first 1,000 word families. Table 4 also shows how the coverage of the BNC/COCA lists drops dramatically from the 1st 1,000 list to the 2nd 1,000 and so on. There are some exceptions to the decrease in coverage, such as the 11th 1,000 because it contains rugby words like *scrum* and *stats*. Note some examples of vocabulary across the levels are given in column 4 to illustrate how this vocabulary is spread throughout high, mid, low and supplementary vocabulary (see Nation, 2016).

Table 4
Lexical profile of spoken rugby corpus across Nation's (2012) BNC/COCA word lists.

BNC/COCA word list	Coverage %	Cumulative coverage %	Examples of vocabulary	
1	84.34	84.34	game, kick	High frequency
2	3.62	87.96	defence, advantage	
3	1.47	89.43	penalty, tackle	
4	0.51	89.94	bonus, prop	Mid-frequency
5	0.32	90.26	referee, intercept	
6	0.28	90.54	rugby, jersey	
7	0.13	90.67	pods, hooker	
8	0.19	90.86	nil, winger	Low frequency
9	0.03	90.89	concussion, lethargic	
10	0.06	90.95	hardcore, maul	
11	0.21	91.16	scrum, stats	
12	0.04	91.20	mongrel, jugular	
13	0.10	91.30	offside, offload	
14	0.10	91.40	ruck, rucks	
15	0.06	91.46	halfback, halfbacks	
16	0.00	91.46		
17	0.10	91.56	lineout, lineouts	
18	0.00	91.56		
19	0.01	91.57	debutant, ropey	
20	0.02	91.59	tighties, tighthead	
Proper nouns	6.39	97.98	Chiefs, Smith	Supplementary lists
Marginal Words	1.63	99.61	fucking, eh, oh	
Transparent compounds	0.26	99.87	outstanding	
Abbreviations	0.07	99.94%	MSP (name of team)	
Additional base word list	0.06	100%	loosehead, loosies	
Total	100			

These lexical profile findings illustrate three important points. First, knowledge of the high frequency word families (1st, 2nd, and 3rd 1,000) is critical for comprehension in spoken rugby discourse. This finding is line with other studies showing that the first 3,000 word families are critical in spoken discourse (Adolphs & Schmitt, 2003; Dang et al., 2017; van Zeeland & Schmitt, 2013). The second important point is that knowledge of proper nouns and marginal words is critical to achieving 95% coverage in spoken rugby. With 8.02% coverage between the four supplementary lists and the base rugby list (see the final five rows of Table 4), these lists provide the next highest coverage of the spoken rugby corpora after high frequency vocabulary. Finally, the number of items in the additional base word list was low, with only 0.06% coverage, which indicates that there only a few possible technical words which are exclusive to rugby, as Ha and Hyland (2017) found in finance. Finally, technical rugby items occur across the frequency levels of the 25,000 BNC/COCA word lists which, as Nation (2016) points out, is a feature of this lexis.

4.2. Research question two: what is the coverage of the single word technical spoken rugby word list over the spoken rugby corpus?

The spoken rugby word list contains 252 types (see Appendix A). The list is organised according to the semantic rating scale from most technical to the items which are shared between everyday and rugby English. As shown in Table 5, 2.7% (seven items) were categorised as highly technical, and roughly a third of the items in the list (33.3%) were categorised as scale 3, meaning they have a rugby-specific meaning which is not the same as their everyday meaning. A total of 161 (63.88%) word types in the spoken word list were categorized as scale 2, which means these words occur in everyday English and in rugby English with the same meaning.

Table 5
Summary of semantic analysis of technical words in the spoken rugby word list.

Semantic scales	Spoken word list (word types)	Coverage %
Semantic rating scale 4: A word that is unique to rugby and only associated with rugby	7	0.41
Semantic rating scale 3: A word that is used in rugby, but with a particular meaning not frequently encountered in everyday usage	84	3.78
Semantic rating scale 2: A word that is used in rugby but with the same meaning most frequently encountered in everyday usage	161	7.85
Total	252	12.04

The results in Table 5 show that the coverage of the word list over the rugby corpus is in line with previous lexical studies, with 12% of vocabulary in spoken discourse being technical (Coxhead & Demecheleer, 2018). This means that roughly one in eight words in spoken rugby discourse analysed in this study is technical. The following excerpt from the TV Commentary corpus highlights the frequency of technical single and MWU rugby vocabulary, as well as proper nouns. The aforementioned items are underlined for ease of recognition.

> Just not making any ground here at all Stirzaker and nice offload here to Dominic Day, ten out from the twenty-two. Tom English goes in and plays halfback, oh, awkward pass and he has knocked it on, I suspect. Ball is loose and it is going to be turned over quickly. Evans hoofs it away down field.
>
> (Grant Nisbett – Highlanders vs. Rebels 31/03/2017)

To provide a sense of the word list in terms of the semantic features of the items, Table 6 presents the ten most frequently occurring technical words in the spoken corpus for each sub-scale by their frequency. All 20 words in the spoken corpus rated as scale 2 and 3 occur in Nation's (2012) BNC/COCA first 1,000 frequency list. This means these words are high frequency both in general spoken discourse and in rugby spoken discourse. The high frequency words *gone* and *behind* in Table 6 are examples of such items. Note that all 10 words rated as scale 4 are low frequency (9,000–25,000), in Nation's (2012) supplementary word lists, or in the additional base word list.

Table 6
Top 10 technical spoken types from semantic rating scale bands 2, 3, and 4 arranged by frequency.

Technical rugby word	Frequency in spoken rugby corpus	Frequency list in Nation's (2012) BNC/COCA, supplementary lists and additional base word list	Semantic rating scale
ball	313		2
out	279		3
back	269		3
work	150		3
game	146		2
kick	135		2
line	130	1st 1000	2
try	118		3
pass	116		2
half	101		3
play	101		3
scrum	94	11th 1000	4
hard	90		3
side	90		3
forward	72		3
gone	69		2
inside	68	1st 1000	3
front	66		2
points	66		2
field	54		2
behind	53		2
lineout	43	17th 1000	4
ruck	34	14th 1000	4
rucks	22	14th 1000	4
lineouts	17	17th 1000	4
scrums	14	11th 1000	4
midfield	14	33rd 1000	4
loosehead	11	35th 1000	4
loosies	10	35th 1000	4
tighties	7	35th 1000	4

4.3. Research question three: What are the technical multiword units in the spoken corpus?

A total of 267 items were identified in the analysis of the spoken corpus (Table 7), and 239 individual root structures were identified (e.g., *set piece, to throw, the pressure*). Of these root structures, 174 are stand-alone (e.g., *the penalty, bonus point, clean out*), meaning they do not have a variable either side of the root structure. The majority of the root structures (154) contain two words, such as *inside pass, weighted kick*, and *ball back*. The variables that occur either side of the root structures are italicised in Table 7.

Table 7
Summary of the technical spoken MWU list (variables in italics).

Information	Number	Example
Root structures	239	set piece, to throw, the pressure
Stand-alone root structures	174	the penalty, bonus point, clean out
Variables	58	*in the* first half, *in from* the side, *back on* the inside
Variables after root structure	27	hands it *off to*, close to *the line*, swings it *away*
Variables before and after root structure	8	*over* the top *of*, *into* the arms *of*, *number* minutes remaining *in the game*
Total number of MWUs	267	

An interesting finding in this analysis is that a single technical word might appear in multiple root structures. Table 8 provides an example from the spoken MWU list for the root structure 'the ball', which has 15 MWUs and 10 root structures. All variables for this MWU occurred before the root structure. The entire spoken MWU list can be seen in Appendix B.

Table 8
Example MWU for the root structure '*the ball*'.

Variable	Root structure
Over	the ball
Off	
Got	
Onto	
Past	
with	
	ball in hand
	ball back
	quick ball
	ball now
	foot ball
	go forward ball
	our ball
	ball is loose
	with ball

One of the key methodological contributions of this paper is the identification of both single and multiword units in rugby. Most word lists in ESP focus on single words. Framing the multiword units around technical single words highlights these items in context and provides chunks of language. The categorisation of these items illustrates the possible variables in technical multiword units and identifies which root structures stand by themselves and which do not.

4.4. Research question four: to what extent might the receptive knowledge of technical rugby vocabulary differ between L1 and L2 rugby speakers?

Table 9 shows the overall results of the receptive knowledge task completed by 77 participants. Out of a total of 15 technical words, the mean score of the L1 speakers (N = 29) was 13 (86.66%) and the mean score of the L2 speakers (N = 48) was 6.6 (44.22%). A Mann–Whitney U test showed no statistical significance between L1 and L2 speakers ($p = 0.522$), which means L1 and L2 speakers' receptive knowledge of technical single words is similar. It may be surprising that the scores in the tasks were not significantly different between the L1 and L2 players. One reason for this finding could be that the participants all had experience with rugby in some capacity.

Table 9
Summary of technical rugby word receptive knowledge task.

Participants	Mean (15)	Standard Deviation	Significance ($p < .05$)
L1 speakers (N = 29)	13 (86.66%)	2.92	$p = 0.522$
L2 speakers (N = 48)	6.6 (44.02%)	4.04	

An analysis of the 15 technical words showed that the L1 speakers had almost full knowledge of the technical words and had a narrow range of scores, with a standard deviation of 2.92. The highest score was 28 of the 29 (96.55%) for an L1 speaker (see Table 10). The L2 speakers had a much wider range of scores, with a standard deviation of 4.04. They also showed high recognition levels of *defence* (40/48; 83.33%), but only 5 (10.41%) recognised *loosies*. This result suggests that there was a wide range of technical vocabulary knowledge among the L2 rugby speakers. Table 10 shows the semantic rating scale and frequency band from the BNC/COCA list (Nation, 2012) for these items. It seems that receptive knowledge was not affected by the semantic rating of the items or the frequency. Two items that stood out from this list to the researchers are *loosies* and *tighties*. With a significance of $p < 0.0001$, it was clear L1 speakers of English knew these items, while L2 speakers did not. One possible reason could be that the items are mainly spoken in the New Zealand rugby context, which further highlights how rugby language can be manifested locally (Kuiper & Lewis, 2013). Table 10 also shows that some words were not known by L1 players (e.g. *scrum*). It could be that participants skipped items or were not paying attention. Future research could involve steps for checking, such as sitting with participants as they complete the task or using post-task interviews.

Table 10
Results of the receptive knowledge task for technical single words.

Word	Semantic rating scale	BNC/COCA level list	L1 (N = 29)	L2 (N = 48)	Significance (p < .0017)
intercept	2	5	28 (96.55%)	17 (35.41%)	p < .0001
winger	3	8	28 (96.55%)	11 (22.91%)	p < .0001
ruck	4	14	28 (96.55%)	29 (60.41%)	p = .0005
halfback	3	15	27 (93.10%)	19 (39.58%)	p < .0001
scrum	4	11	27 (93.10%)	35 (72.91%)	p = .0302
lineout	4	17	27 (93.10%)	26 (54.16%)	p = .0003
defence	2	2	26 (89.65%)	40 (83.33%)	p = .236
penalty	2	3	26 (89.65%)	30 (62.5%)	p = .004
back	2	2	26 (89.65%)	16 (33.33%)	p < .0001
offload	3	13	25 (86.20%)	33 (68.75%)	p = .085
loosies	4	35	24 (82.75%)	5 (10.41%)	p < .0001
momentum	2	4	22 (75.86%)	14 (29.16%)	p < .0001
jersey	2	6	22 (75.86%)	8 (16.66%)	p < .0001
breakdown	3	33	21 (72.41%)	28 (58.3%)	p < .0011
tighties	4	20	19 (65.51%)	6 (12.5%)	p < .0001

5. Pedagogical implications

There are several implications for players, coaches and English for Specific Purposes teachers based on the findings of this study. These implications might also extend to course development, goal setting, and materials development. Table 11 focuses on the knowledge of technical rugby vocabulary that L2 players need in rugby based on the findings of this study and what actions could be taken to help with developing that knowledge by learners themselves, language teachers and coaches. The pedagogical suggestions in Table 11 are not in order of priority, as the results of the receptive knowledge task suggest knowledge is not affected by the semantic rating of the items. However, ESP teachers should take technical vocabulary knowledge into account when creating such a curriculum.

Table 11
Knowledge of technical vocabulary needed by L2 players and suggestions for learning and teaching.

What technical vocabulary knowledge do L2 players need?	What can learners and teachers do?
Highly technical words in the word list that are unique to rugby and only associated with rugby (see Appendix A)	Undertake direct learning through word cards. Use repetition to ensure there are plenty of encounters. Make sure that these words are recognised in written and spoken form, and used fluently in speaking.
Words that are used in rugby with a particular meaning not frequently encountered in everyday usage (e.g. *side* meaning *team* in rugby).	Use strategies such as word cards, seeking exposure to these words in the rugby community, for example, by playing the game or watching English TV commentary, and talking about rugby. Reading rugby magazine articles and websites would also be useful.
Words in the technical word list of rugby that are used in rugby but with the same meaning most frequently encountered in everyday usage	Work with the technical multiword list to learn common patterns of use in rugby that might be different from general English. Ensure there are plenty of opportunities to encounter these words in rugby contexts through conversations, reading magazine and websites, and watching rugby on TV.
Technical multiword units of rugby	Use the list as a reference for how words are used in common patterns in rugby, focus on word combinations for high frequency items in the single word list, such as *ball*, and take opportunities to use the multiword units in speaking about rugby, at practice, and during games.
Technical multiword units stand-alone structures	Learn these combinations as a chunk using word cards.
Technical multiword units with variables	Take note of common variables in multiword units and practise using them in speaking about rugby in preparation for playing or practising the game. Also listen for them in TV commentary.
'Local' terms and swear words, e.g., *loosies*	Learn these items directly in preparation for playing in a local context.

Note that high frequency vocabulary is needed for understanding rugby talk. To focus on learning this vocabulary, learners could use word cards and language programmes could ensure there are plenty of graded reading and listening materials as well as opportunities to talk and write using this vocabulary.

5.1. Limitations

A limitation of the study is the size of the spoken rugby corpus. As noted in the methodology section, automatic transcription was not possible and therefore, the TV Commentary corpus had to be manually transcribed. This was also an issue for Wilson (2011) when creating the Interactional Corpus. The size of the corpora affects the generalisability of the study and

using range as an additional selection principle for the technical spoken rugby word list. Another limitation concerns the raters used during the semantic rating analysis. Although three raters allowed for a decision to be made if there were conflicting results, more raters would have been beneficial to provide concrete results on the rating of technical items and allow for statistical analysis of the data. Finally, the technical word lists are limited to the selection of New Zealand rugby discourse in the corpus and, therefore, cannot be representative of technical rugby words in different contexts.

5.2. Future research

As this is the first study to investigate the lexicon of spoken rugby discourse, much more research can be done in this area. Future research could look at other types of discourse in rugby, such as gathering more authentic communication in areas that Wilson (2011) was unable to access, such as during a game. Another type of discourse could include commentaries from radio. Kuiper and Lewis' (2013) study investigated speech between radio and TV rugby commentary, with their results showing radio provides more detailed commentary than TV. Further investigations could include written texts such as rule books. A lexical profile analysis may provide more occurrences of technical rugby vocabulary. Finally, investigating language used during the game by other members of the rugby community, such as the referee, could highlight more important vocabulary necessary for second language learners to successfully assimilate into the foreign rugby setting. To date, only one study has investigated referee talk (Vine, 2017).

Another area of interest could be to explore technical rugby vocabulary in different countries. This could entail creating similar corpora in other rugby countries, such as Japan or England and comparing the technical vocabulary, so as to assist rugby players and coaches wanting to join those communities. An anonymous reviewer suggested a comparison with other languages, which could perhaps result in a bilingual word list such as the Tongan/English trades education lists by Coxhead et al. (2020b). Prior to creating more word lists in other spoken rugby contexts, future research on validating the created technical rugby word lists in this study is needed.

6. Conclusion

The study has shed light on a previously unexplored area of ESP through an investigation on the lexical demands of spoken rugby discourse. It is now understood learners will need to know the first 4,000 word families of Nation's (2012) BNC/COCA 25,000 and supplementary lists to comprehend rugby discourse. Furthermore, knowledge of proper nouns is critical to understanding TV rugby commentary and marginal words, such as swear words, are crucial when interacting in a rugby team. The development of rugby single-word and multiword technical word lists provide insights into the nature of technical vocabulary in non-university ESP domains, and are valuable resources in the design of a specialised curriculum to meet the learners' lexical needs. In addition, it provides valuable contributions to the creation of specialised spoken corpora outside of traditional settings.

Appendix A. Technical spoken rugby word list in order of semantic rating

Semantic rating scale 4 (words are unique to rugby and are associated with rugby)

1. scrum - s
2. lineout - s
3. ruck - s
4. Midfield
5. loosehead
6. loosies
7. tighties

Semantic rating scale 3 (words occur in rugby but with a different meaning to that of everyday usage)

1.	out	43.	shot
2.	back	44.	prop
3.	work - ing	45.	flat
4.	try - ies	46.	drive
5.	play - s	47.	pick
6.	half	48.	picked
7.	hard	49.	touch
8.	side	50.	holding
9.	hit - s, ting	51.	wing
10.	run - ning	52.	winger
11.	forward	53.	center
12.	inside	54.	dropped
13.	hands	55.	break
14.	set	56.	cut
15.	man	57.	row
16.	number - s	58.	build

17. outside
18. advantage
19. forwards
20. hold
21. space
22. phase - s
23. backs
24. loose
25. rugby
26. call
27. possession
28. season
29. lost
30. wide
31. takes
32. close
33. halfback
34. clean
35. knocked
36. knock
37. tight
38. breakdown - s
39. fullback
40. feet
41. mark
42. round
59. form
60. bench
61. conversion
62. full
63. pack
64. Posts
65. room
66. sides
67. breaks
68. clear
69. hole
70. read
71. turnover
72. called
73. fires
74. free
75. pops
76. sweet
77. gate
78. ref
79. calling
80. carried
81. class
82. deck
83. drill
84. openside

Semantic rating scale 2 (Words occur in the rugby context but also have the same meaning in everyday usage)

1. ball
2. kick - ing, s, ed
3. game - s
4. pass - es, ing
5. line
6. point - s
7. penalty - ies
8. player - s
9. tackle - s
10. playing - ed
11. gone
12. defence
13. front
14. team - s
15. field
16. behind
17. opportunity - ties
18. taken
19. meter - s
20. push - ed
21. throw - s, n
22. move - s
23. ground
24. quick
25. moment
26. short
27. position
28. across
29. talk - ing
30. attack
31. pressure
32. defender - s
33. fifteen
34. win - won
35. charge - ing
36. option - s
37. offside
38. twelve
39. goal
40. long
41. piece
42. offload - s
43. high
44. step
45. match
82. training
83. whistle
84. focus
85. low
86. progress
87. zone
88. momentum
89. shoulder
90. width
91. boot
92. decision
93. distance
94. result
95. caught
96. dangerous
97. doubt
98. easy
99. moving
100. rate
101. stand
102. standing
103. support
104. keen
105. replacement
106. scores
107. scoring
108. solid
109. standard
110. pods
111. backwards
112. depth
113. fight
114. final
115. missed
116. attacking
117. create
118. impressive
119. replaced
120. swings
121. defending
122. stats
123. ahead
124. club
125. feed
126. force

46.	middle	127.	lose
47.	score	128.	pull
48.	job	129.	battle
49.	speed	130.	bounce
50.	bonus	131.	claimed
51.	arm - s	132.	individually
52.	pace	133.	injuries
53.	referee	134.	lengths
54.	contact	135.	opposition
55.	territory	136.	progression
56.	outstanding	137.	tough
57.	held	138.	defend
58.	lead - ing	139.	errors
59.	grab - bed	140.	setup
60.	took	141.	catch
61.	top	142.	chances
62.	patterns	143.	corners
63.	gap	144.	danger
64.	halftime	145.	face
65.	deep	146.	fielded
66.	defensive	147.	late
67.	percent	148.	mistakes
68.	remaining	149.	showed
69.	scored	150.	shut
70.	beat	151.	used
71.	place	152.	captain
72.	defensively	153.	directly
73.	positive	154.	excellent
74.	dominate	155.	injured
75.	drop - s	156.	intensity
76.	finish	157.	senior
77.	power	158.	shove
78.	snapped	159.	smash
79.	angle	160.	split
80.	jersey	161.	intercept
81.	nil		

Appendix B. Technical spoken MWU list in order of semantic rating scale and the frequency of the technical single item in the semantic rating scale. The technical single item is italicized

Semantic rating scale 4

Number	Variable	Root structure	Variable
1	In	behind the *scrum*	
2		a *scrum*	
3		good *scrum*	
4		first *scrum*	
5		*scrum* time	
6		the *ruck*	
7		a *ruck*	
8		at *ruck*	
9		the *lineout*	
10		a *lineout*	
11		in *midfield*	

Semantic rating scale 3

Number	Variable	Root structure	Variable
1	go	*out* there	
2	get	*out* there	
3	to	get *out*	of
4		*out* on	
5		*out* the back	
6		clean *out*	
7		*out* here	
8		*out* in	

Appendix B. (*continued*)

Number	Variable	Root structure	Variable
9		*out* of	there
10		*out* of	this
11		*out* of	it
12		*out* wide	
13		right *out*	
14		coming *out*	
15		is *out*	
16		*out* from	
17	on	the *back*	
18	out	the *back*	
19	at	the *back*	
20		the *back*	of
21	on	the *back*	of
22	out	the *back*	of
23	at	the *back*	of
24		*back* to	
25		*back* in	
26		*back* on	the inside
27		come *back*	
28		*back* line	
29		get *back*	
30	(name of player)	is *back*	
31		*back* inside	
32	to	go *back*	
33		*back* into	
34	the	ball *back*	
35		*back* there	
36		going *back*	
37		*back* by	
38		*back* up	
39		got *back*	
40	the	*try* line	
41		to *try*	
42		the *try*	
43		a *try*	
44		*try* to	
45		good *work*	
46		*work* hard	
47	got	to *work*	
48		the *work*	of
49		*work* on	
50		*work* rate	
51		hard *work*	
52	in	first *half*	
53	in the	first *half*	
54	in the	second *half*	
55		*half* way	
56		this *half*	
57		to *play*	
58		let's *play*	
59		*play* on	
60		work *hard*	
61		really *hard*	
62		*hard* work	
63		working *hard*	
64	from	the *side*	
65	in from	the *side*	
66		that *side*	
67		this *side*	
68	on the	far *side*	
69	on the	other *side*	
70		wing *side*	
71		go *forward*	ball
72		lost *forward*	

(*continued on next page*)

Appendix B. (*continued*)

Number	Variable	Root structure	Variable
73		*inside* the	number
74	back on	the *inside*	
75	on	the *inside*	
76		back *inside*	
77		*inside* pass	
78		*hands* it	off
79		*hands* it	off to
80		*hands* it	on
81		quick *hands*	
82		*hands* on	
83		*hands* up	
84		his *hands*	
85		the *hands*	
86		*set* piece	
87		*set* up	
88		a *set*	
89		to *set*	
90		the *man*	
91		*man* down	
92		to *run*	
93	on	the *outside*	
94		the *forwards*	
95	a	bit of *space*	
96	in a	bit of *space*	
97		some *space*	
98		to *hold*	
99		*hold* it	
100		*hold* onto	
101	number	*tries* to	name
102		out *wide*	
103		go *wide*	
104		*takes* it	in
105		*close* to	the line
106		*knocked* on	by
107	a	*knock* on	
108		his *feet*	
109		*pick* up	
110		*picked* up	
111		have *picked*	
112		the *full*	
113	under	the *posts*	
114		both *sides*	
115		*sweet* as	
116	over the	*advantage* line	
117		an *advantage*	
118		no *advantage*	
119		ball is *loose*	
120		*loose* pass	
121		the *season*	
122		front *row*	
123	through	the *gate*	
124		the *ref*	
125		the *bench*	

Appendix B. (*continued*)

Number	Variable	Root structure	Variable
126		at *halfback*	
127		the *halfback*	

Semantic rating scale 2.

Number	Variable	Root structure	Variable
1	in	the *game*	
2	in	this *game*	
3	over	the *ball*	
4	off	the *ball*	
5	got	the *ball*	
6	onto	the *ball*	
7	past	the *ball*	
8	with	the *ball*	
9		*ball* in hand	
10		*ball* back	
11		quick *ball*	
12		*ball* now	
13		foot *ball*	
14		go forward *ball*	
15		our *ball*	
16		*ball* is loose	
17		with *ball*	
18	the	*kick* off	
19	with	the *kick*	
20		good *kick*	
21		*kick* it	
22		to *kick*	
23		*kick* in	
24		little *kick*	
25		weighted *kick*	
26	to	the *line*	
27	on	the *line*	
28	close to	the *line*	
29	the	advantage *line*	
30	over the	advantage *line*	
31		back *line*	
32		defensive *line*	
33		*line* speed	
34		the side *line*	
35	number	meter *line*	
36	the	try *line*	
37		has *gone*	
38		have *gone*	
39		minutes *gone*	
40		in *front*	of
41		up *front*	
42		the *front*	
43		*front* row	
44		*front* foot	
45		*front* foot ball	
46		the *pass*	
47		short *pass*	
48		loose *pass*	
49		*pass* away	
50		*pass* off	
51		inside *pass*	
52		*points* to	number
53	On	the *field*	
54	away	down *field*	
55		in *behind*	the scrum
56		*taken* by	(name of player)
57		*taken* down	by

(*continued on next page*)

Appendix B. (*continued*)

Number	Variable	Root structure	Variable
58		*taken* in	
59		nicely *taken*	
60	on	the *ground*	
61		to *ground*	
62		*quick* hands	
63		*quick* ball	
64		*push* ups	
65		*push* it	
66		to *push*	
67	as	a *team*	
68	at	the *moment*	
69		the *moment*	anyway
70		bonus *point*	
71		*short* pass	
72	name	goes *short*	
73		*move* it	
74		to *move*	
75		set *piece*	
76	down	the *middle*	
77	in	the *middle*	
78		the *middle*	of
79	into	the *arms*	of
80		*talk* about	
81		*talk* to	him
82		to *throw*	
83		*held* up	
84		not *held*	
85	over	the *top*	
86	over	the *top*	of
87		on *top*	
88		to *win*	
89		the *charge*	
90		no *doubt*	
91		work *rate*	
92		the *club*	
93		*fielded* by	(name of player)
94		on *defence*	
95		good *defence*	
96		in *defence*	
97		the *defence*	
98		(name of team) *defence*	
99		their *defence*	
100		to *attack*	
101		*pressure* on	
102		the *pressure*	
103		under *pressure*	
104		the *goal*	
105		the *match*	
106		good *option*	
107		the *patterns*	
108		*defensive* line	
109	number	minutes *remaining*	in the game

Appendix B. (*continued*)

Number	Variable	Root structure	Variable
110		*snapped* up	by
111		the *replacement*	
112		being *replaced*	
113		*swings* it	away
114		*claimed* by	(name of player)
115		the *bounce*	
116		the *captain*	
117	Be	a *penalty*	
118	got	a *penalty*	
119		the *penalty*	
120		another *penalty*	
121	in	the *tackle*	
122	a	good *tackle*	
123		*tackle* by	(name of player)
124		the *gap*	
125		to *dominate*	
126		*bonus* point	
127		the *referee*	
128	number	to *nil*	
129		at *halftime*	

References

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24*(4), 425-438.

Aiguo, W. (2007). Teaching aviation English in the Chinese context: Developing ESP theory in a non-English speaking country. *English for Specific Purposes, 26*, 121-128.

Basturkmen, H. (2010). *Developing courses in English for specific purposes*. Palgrave Macmillan.

Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253-279.

Bergh, G. (2011). Football is war: A case study of minute-by-minute football commentary. *Veredas, 2*, 83-93.

Bergh, G., & Ohlander, S. (2012). English direct loans in European football lexis. In C. Furiassi (Ed.), *Anglicization of European lexis* (pp. 281-304). John Benjamins.

Bergh, G., & Ohlander, S. (2017). Loan translation versus direct loans: The impact of English on European football lexis. *Nordic Journal of Linguistics, 40*(1), 5-35.

Browne, K. (2005). Snowball sampling: Using social networks to research non-hetrosexual women. *International Journal of Social Research Methodology, 8*(1), 47-60.

Chiba, N., & Jackson, S. (2006). Rugby player migration from New Zealand to Japan. *Football Studies, 9*(2), 67-78.

Chung, T. M., & Nation, I. S. P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language, 15*(2), 103-116.

Chung, T. M., & Nation, I. S. P. (2004). Identifying technical vocabulary. *System, 32*(2), 251-263.

Coxhead, A. (2000). A new academic word list. *Tesol Quarterly, 34*(2), 213-238.

Coxhead, A. (2018). *Vocabulary and English for specific purposes research: Quantitative and qualitative perspectives*. Routledge.

Coxhead, A. (2019). Working with multi-word units in ESP/EAP. In E. Hinkel (Ed.), *Teaching essential units of language: Beyond single-word vocabulary* (pp. 36-54). Routledge.

Coxhead, A., & Demecheleer, M. (2018). Investigating the technical vocabulary of Plumbing. *English for Specific Purposes, 51*, 84-97.

Coxhead, A., Demecheleer, M., & McLaughlin, E. (2016). The technical vocabulary of carpentry: Leads, lists and bearings. *TESOL Journal, 25*, 38-71.

Coxhead, A., McLaughlin, E., & Reid, A. (2018). The development and application of a specialised word list: The case of Fabrication. *Journal of Vocational Education and Training, 71*(2), 175-200.

Coxhead, A., Parkinson, J., Mackay, J., & McLaughlin, E. (2020). *English for vocational purposes: Language use in trades education*. Routledge.

Coxhead, A., Parkinson, J., & Tu'amoheloa, F. (2020). Using Talanoa to develop bilingual word lists of technical vocabulary in the trades. *International Journal of Bilingual Education and Bilingualism, 23*(5), 513-533.

Dang, T. N. Y. (2018a). The hard science spoken word list. *ITL - International Journal of Applied Linguistics, 169*(1), 44-71.

Dang, T. N. Y. (2018b). The nature of vocabulary in academic speech of hard and soft- sciences. *English for Specific Purposes, 51*, 69-83.

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning, 67*(4), 959-997.

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes, 33*, 66-76.

Desmarais, F., & Bruce, T. (2009). The power of the local in sports broadcasting: A cross-cultural analysis of rugby commentary. *International Journal of Sport Communication, 2*, 129-145.

Desmarais, F., & Bruce, T. (2010). The power of stereotypes: Anchoring images through language in live sports broadcasts. *Journal of Language and Social Psychology, 29*(3), 338-362.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes, 28*, 157-169.

Fox, J., & Tigchelaar, M. (2015). Creating and engineering academic formulas list. *Journal of Teaching English for Specific and Academic Purposes, 3*(2), 295-304.

Görlach, M. (2001). *A dictionary of European Anglicisms*. Oxford University Press.

Ha, Y.-H., & Hyland, K. (2017). *What is technicality? A technicality model for EAP vocabulary* (vol. 28, pp. 35-49)English for Academic Purposes.

Harris, T. (2003). Listening with your eyes: The importance of speech-related gestures in the language classroom. *Foreign Language Annals, 36*(2), 180-187.

Harris, J. (2010). *Rugby union and globalization: An odd-shaped world*. Palgrave.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). *RANGE and FREQUENCY programs*. Retrieved from https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs.

Humpolík, R. (2014). *Language of football commentators: An analysis of live English football commentary and its types* (Unpublished Bachelor's Diploma Thesis). Masaryk University.
Kuiper, K. (1996). *Smooth talkers: The linguistic performance of auctioneers and sportscasters*. Lawrence Erlabum.
Kuiper, K., & Lewis, R. (2013). The effect of the broadcast medium on the language of radio and television sports commentary genres: The rugby union lineout. *Journal of Sports Media, 8*(2), 31-51.
Lavric, E., Pisek, G., Skinner, A., & Stadler, W. (2008). *The linguistics of football*. Gunter Narr.
Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59-82.
Nation, I. S. P. (2012). *The BNC/COCA word family lists*. Retrieved from https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists.
Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.
Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins.
Parkinson, J., & Mackay, J. (2016). The literacy practices of vocational training in Carpentary and Automotive Technology. *Journal of Vocational Education and Training, 68*(1), 33-50.
Quero, B. (2015). *Estimating the vocabulary size of L1 Spanish ESP learners and the vocabulary load of medical textbooks* (Unpublished doctoral dissertation). Victoria University of Wellington.
Schmidt, T. (2008). The Kictionary: Combining corpus linguistics and lexical semantics for a multilingual football dictionary. In E. Lavric, G. Pisek, A. Skinner, & W. Stadler (Eds.), *The linguistics of football*. Gunter Narr.
van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457-479.
Vine, E. W. (2017). Contructing coopoerative and antagonistic intercorporeality: Referee talk and action on the field. In C. Meyer, & U. V. Wedelstaedt (Eds.), *Moving bodies in Interaction - interacting bodies in motion: Intercoporeality, interkinesthesia, and enaction in sports* (pp. 113-151). John Benjamins.
Watson-Todd, R. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes, 45*, 31-39.
Webb, S., & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics, 30*(3), 407-427.
Webb, S., & Rodgers, M. P. H. (2009b). Vocabulary demands of televison programs. *Language Learning, 59*(2), 335-366.
Wilson, N. (2009a). The discourse of deputies: Communicating co-leadership in a rugby club. *Te Reo, 52*, 73-98.
Wilson, N. (2009b). "Fucking" as a solidarity marker. In *Paper presented at the the Linguistic Society of New Zealand, Massay University*.
Wilson, N. (2011). *Leadership as communicative practice: The discoursive construction of leadership and team identity in a New Zealand rugby team* (Unpublished doctoral dissertation). Victoria University of Wellington.
Wood, D. C., & Appel, R. (2014). Multiword constructions in first year business and engineering university textbooks and EAP textbooks. *Journal of English for Academic Purposes, 15*, 1-13.
Woodward-Kron, R. (2008). More than just jargon: The nature and roles of specialist knowledge in learning disciplinary knowledge. *Journal of English for Academic Purposes, 7*, 234-249.
Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Stuart Benson is an Associate Professor at the University of Aizu in Fukushima, Japan. He has taught in Japan and New Zealand. His areas of interest are ESP in non-university settings, vocabulary acquisition, and corpus linguistics. His current research investigates technical vocabulary in spoken rugby discourse.

Averil Coxhead is a Professor of Applied Linguistics in the School of Linguistics and Applied Language Studies, Victoria University of Wellington, New Zealand. Her current research interests include technical vocabulary in trades education, measuring knowledge of technical vocabulary, and vocabulary in English for Specific Purposes and English for Academic Purposes.