**LEARN Journal: Language Education and Acquisition Research Network**
ISSN: 2630-0672 (Print) | ISSN: 2672-9431 (Online)
Volume: 18, No: 2, July – December 2025

Language Institute, Thammasat University
https://so04.tci-thaijo.org/index.php/LEARN/index

# "*Dead at the Scene*" and More: Lexical Bundles in Accident News

**Mathurin Leelasetakul**

Mathurin.L@Chula.ac.th, Department of English, Chulalongkorn University, Thailand

**APA Citation:**
Leelasetakul, M. (2025). "Dead at the scene" and more: Lexical bundles in accident news. *LEARN Journal: Language Education and Acquisition Research Network, 18*(2), 78-107. https://doi.org/10.70730/EUQT4958

Received
01/12/2024

Received in revised form
15/02/2025

Accepted
22/02/2025

**ABSTRACT**

This study explores lexical bundles in vehicular accident news with the goal to provide a lexical bundle list for students learning to write this genre of news in English. A corpus of accident news is constructed from vehicular news articles from four English news sources over a period of one year. The lexical bundles are extracted from using the frequency-based approach in which the frequency cut-off point for this study is at 60 per million words, and they must also appear at least in three different texts and in two subcorpora. In total, there are 79 lexical bundles identified for vehicular accident news and they are further categorized based on their functions as stance expressions, discourse organizers, referential expressions, and news-specific functions. The patterns and contexts of usage for each lexical bundles are discussed in detail so that they can be of use for teachers and students.

**Keywords:** corpus linguistics, lexical bundles, formulaic language

## Introduction

This present study arises from a pedagogical need to equip students with some assistance when it comes to translating a specific genre of news, vehicular accident news, from Thai to English. One struggle that many students face is not having enough input on how to formulate sentences for news in English due to the fact that they might not have read a lot of or might not be familiar with news in English. Thus, if such formulaic expressions as lexical bundles, which are "sequences of words that commonly go together in natural discourse" (Biber & Conrad, 1999, p. 185), can function as "building blocks" (Biber & Conrad, 1999, p. 185) for language learners when writing in their second language, having a list of frequently occurring lexical bundles in this genre of English news or having class materials that were developed based on such a list can help students tremendously as they navigate how to translate and write accident news in English.

## Literature Review

### Formulaic Language

Language in use, both in speech and in writing, consists of a large proportion of what is called formulaic sequences (Nattinger & DeCarrico, 1992; Schmitt & Carter, 2004). This structure is defined by Wray (2002) as "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use" (p. 465). Wray (2002) has also compiled an extensive list of terms that have been used to refer to this phenomenon, ranging from "collocations", "sentence builders", "lexical or lexicalized phrases", "multiword units" to "stock utterances". Moon (1997) calls this language phenomena "multi-word items" (p. 40). Schmitt and Carter (2004) also highlight frequency as one of the key characteristics of formulaic language.

These language units are processed much quicker in reading tasks than non-formulaic ones in both native speakers and in second language learners as seen from the reaction times in experiments because such formulaic units are "processed holistically as single lexicalized units" (Jiang & Nekrasova, 2007, p. 441). In terms of production, several studies on writing have shown that students benefited from instruction of formulaic sequences as can been seen in post-test scores and actual usage of formulaic sequences after instruction (Peters & Pauwels, 2015); in addition, students also are found to rely on the instructed formulaic sequences in writing tasks (AlHassan &

Wood, 2015). Hence, this feature of language is important in second language acquisition and instruction, and it is important for second language learners to master the use of these formulaic sequences (Nattinger & DeCarrico, 1992; Wray, 2002).

## Lexical Bundles

The aspect of formulaic sequences that is of interest in this present study is multi-word expressions called lexical bundles. Lexical bundles are defined as "sequences of words that commonly go together in natural discourse" in either conversation or writing (Biber & Conrad, 1999, p. 184) and, in most lexical bundle studies, lexical bundles are four words in length, "identified empirically" (Cortes, 2004, p. 400) and extracted from the corpus using the frequency-based approach.

Lexical bundle studies are "frequency-driven" (Biber & Barbieri, 2007, p. 267) in which frequent word combinations are selected for analysis. The frequency cut-off points (normalized frequency per million words) would depend on the researchers and are to a certain degree arbitrary; for example, Hyland (2008a) and Cortes (2004) used the frequency of 20 per million, whereas Biber et al. (2004) used the frequency of 40 per million words.

In addition to the frequency cut-off points, there are further criteria to include or exclude sequences of words that have been extracted from a corpus in the analysis. For example, for Biber and Conrad (1999), to be considered lexical bundles, the sequence must occur uninterrupted together as a unit and does not cross the sentence boundary. Furthermore, they must occur in multiple texts and must not be produced by a single participant to avoid inclusion of frequent phrases from just one person (Biber & Barbieri, 2007). An example of this criteria is the study by Biber and Conrad (1999), in which a lexical bundle must occur in 5 different texts in the corpus. In certain studies, word sequences that overlap with each other, either as a "complete overlap" or as a "subsumption" (Chen & Baker, 2010, p. 33), need to be merged. Another optional criterion is excluding discipline specific phrases or proper nouns, or analyzing them separately from the more general lexical bundles.

Lexical bundles are usually categorized in terms of structure and function. Structurally, they are usually grouped according to the grammatical structures that they contain. In their study on lexical bundles in textbooks and lectures, Biber et al. (2004) proposed a taxonomy that categorized lexical bundles into three broad groups: those that incorporate (1) verb phrase fragments, (2) dependent clause fragments, and (3) noun phrase and prepositional phrase fragments. Chen and Baker (2010), in their comparison

of native and non-native speaker bundles, grouped lexical bundles slightly differently into three groups: verb-phrase-based, noun-phrase-based and prepositional-phrase-based.

Functionally, lexical bundles can be grouped primarily into three categories: stance expressions, discourse organizers and referential expressions (Ädel & Erman, 2012; Biber et al., 2004; Chen & Baker, 2010; Simpson-Vlach & Ellis, 2010), although different studies may have different functional categories depending on the genre of the corpus. For example, in the academic writing genre, Hyland (2008b) categorized the lexical bundles in his analysis as "research-oriented," "text-oriented" and "participant-oriented" (p. 42). For this study, since the lexical bundles are in the genre of the accident news, the broader functional categories by (Biber et al., 2004; Simpson-Vlach & Ellis, 2010) will be adopted to categorize the lexical bundles extracted from the corpus.

## Word and Phrase Lists

Providing wordlists for second language learners is by no means a new concept. English language learners may be familiar with the Oxford 3000 or the Oxford 5000, which are lists of 3,000 and 5,000 frequently used English words that all and advanced learners, respectively, should be familiar with. Another word list that is a valuable resource in English language learning and teaching with respect to academic writing is the Academic Word List (AWL) compiled by Coxhead (2000), with its subsequent word lists for a specific academic register and academic discipline, which are the Academic Spoken Word List (Dang et al., 2017) and the EAP Science List (Coxhead & Hirsh, 2007).

There also exist studies that investigate lists of lexical items that are longer than one word. One example is a list for frequent collocations in academic writing, the Academic Collocation List (ACL) by Ackermann and Chen (2013). Another prime example is the Academic Formula List (AFL) by Simpson-Vlach and Ellis (2010) with frequent lexical bundles found across different fields of study, both in speech and in writing. With this list, Simpson-Vlach and Ellis (2010) state that it is meant to be used as a "starting point" for subsequent research and for further teaching materials development (p. 502), rather than using the list 'as is' in teaching.

In recent years, there have been more studies that focus on lexical bundles in specific genres or disciplines, from lexical bundles in informal business emails (Siricharoen & Wijitsopon, 2020), to lexical bundles in health news (Xu & Sun, 2022), and to lexical bundles in travel vlogs (Laosrirattanachai & Laosrirattanachai, 2024). These, in turn, have inspired this current study to create a lexical bundle list in the genre of vehicular

accident news to be of further use when designing classroom activities and materials and as a potential resource for students. In this research, as the focus is on how the lexical bundles are used in the news, the focus of the analysis is on their functions.

## Methodology

The corpus for this study is constructed from accident news from four English online news sources that was reported over the period of one year. The reason behind collecting the data for the whole year is to ensure that accident news that is more frequent during a certain time of the year, e.g. during summer or the holidays season when people tend to travel more often, are included.

The types of news that comprise the corpus are vehicular accident news, such as car crashes, train collisions, aircraft crashes and maritime vehicle capsizing. In this aspect, the data collection is controlled for content with the focus on the report of the accidents themselves and would exclude reports on, for example, criminal trials and lawsuits that happen after accidents. Moreover, incidents that happened on such modes of transport but were not accidents were also excluded. An example of such a case is a news article about a man who stabbed fellow passengers and subsequently set fire to the vehicle. Finally, tweets or images with text, such as an infographic, that were embedded in the news are not included although the quotes from other sources in the news text are preserved but are not included in the analysis if a lexical bundle straddles the text in the news and the quoted speech. The details on the number of texts, word types, and tokens (or word count) for each of the four subcorpora are presented in Table 1 below.

**Table 1**

*Accident News Corpus Size*

| Corpora | Number of texts | Type | Word count (tokens) |
|---|---|---|---|
| News subcorpus 1 | 121 | 4,524 | 26,722 |
| News subcorpus 2 | 82 | 5,391 | 27,424 |
| News subcorpus 3 | 76 | 4,456 | 22,520 |
| News subcorpus 4 | 82 | 5,252 | 32,541 |
| **Total** | **361** | **19,623** | **109,207** |

From the table above, in total there are 361 texts in this corpus of accident news and the size of the corpus is 109,207 words. The size of the corpus may appear relatively small when compared with large-scale corpora

such as COCA; however, as the purpose of the analysis is for a specific genre of writing, a corpus of this size is sufficient (Flowerdew, 2004) and can provide "reliable results" (Koester, 2010, p. 68). The goal of this study is not to provide an exhaustive list of just the lexical bundles, but rather to explore how frequent lexical bundles are used in context in a way that will be meaningful for students.

For the analysis of the lexical bundles, the program AntConc (version 4.3.0) is used to extract four-word n-grams from the corpus. The frequency-based approach is adopted to ensure that all frequent lexical bundles in this news genre are included in the analysis. The frequency cut-off point for the n-grams to be included in the analysis is at 60 times per million. Furthermore, for range, in order to be included in the list, a lexical bundle must appear in at least three different texts and in at least two subcorpora. The final inclusion criterion is that the lexical bundle needs to occur in more than one accident. This measure is put in place, based on Biber's (2009) criterion to "guard against idiosyncratic uses by individual speakers or authors" (p. 282), to prevent inflation of frequent bundles or proper nouns from just one incident. For example, the phrase "the Mexico City Metro" occurs more than 60 times per million but it is used only in news articles about one metro accident in Mexico.

## Results and Discussion

After filtering out the lexical bundles extracted from the corpus according to the exclusion criteria in the previous section, 79 lexical bundles are identified for vehicular accident news.

The 79 lexical bundles extracted from the corpus are further categorized into their functional categories. The categories are based on the ones used by Biber et al. (2004) and Simpson-Vlach & Ellis (2010), which are stance expressions, discourse organizers and referential expressions. One more category, news-specific functions, is added to include lexical bundles whose functions are clearly specific to the genre of accident news. In total, the lexical bundles can be grouped into four main categories as seen in Table 2 below.

**Table 2**

*Number of Lexical Bundles in Accident News in Each Functional Category*

| Functional Category | Number of lexical bundles | Percentage |
|---|---|---|
| I. Stance Expressions | 2 | 2.5% |
| II. Discourse organizers | 1 | 1.5% |

| | | |
|---|---|---|
| III. Referential Expressions | 29 | 37% |
| V. News-Specific Functions | 47 | 59% |
| Total | 79 | 100% |

As seen from the table, lexical bundles that perform news-specific functions make up the major category at 59% of the total bundles identified, followed by referential lexical bundles at 37%. The high proportions of these two functional categories can be attributed to the nature of the corpus content that is highly informative and genre specific. In contrast, stance expression and discourse organizing lexical bundles combined only makes up 4% of the lexical bundles extracted from the corpus. The lexical bundles that fall into each main functional category and their sub-categories are listed in Table 3 below.

**Table 3**

*Functions of Lexical Bundles in Accident News*

| | |
|---|---|
| **I. Stance Expressions** | *was believed to have* *what appeared to be* |
| **II. Discourse Organizers** | *as well as the* |
| **III. Referential Expressions** | A. Identification/focus<br>*one of the worst* *one of the deadliest* *one of the most*<br>*is one of the* *were among the dead*<br>*children were among the*<br><br>B. Specification of Attributes<br><br>B1) Quantity Specification<br>*more than 100 people* *at least three people*<br>*of the passengers were* *accidents are common in*<br>*are not uncommon in*<br><br>B2) Tangible framing attributes<br>*the body of a* *the driver of the*<br>*the scene of the* *the site of the*<br><br>C. Time/place reference<br><br>C1) Place Reference<br>*near the town of* *near the city of* *near the village of*<br>*south of the capital* *off the coast of*<br>*on a highway in* *a car crash in* *at the junction of*<br>*in the middle of* *in the opposite direction*<br>*were on board the* |

| | |
|---|---|
| | C2) Time Reference<br>*at the time of* *in the early hours*<br><br>C3) Multifunctional Reference<br>*at the end of* |
| **IV. News-Specific Functions** | A. Summary of Accidents<br><br>A1) Summary of Accidents (General)<br>*in a collision with* *was involved in a*<br>*been involved in a*<br><br>A2) Summary of Accidents (Death)<br>*people were killed when* *people were killed in*<br>*people were killed and* *people have been killed*<br>*that killed at least* *has died in a*<br>*people have died in* *died in the crash*<br>*died in a crash*<br><br>B. Details of Accidents<br><br>B1) Details of Accidents (General)<br>*the accident took place* *was hit by a*<br>*crashed into the sea* *and burst into flames*<br>*they were travelling in* *as they tried to*<br>*into the rear of* *off the road and*<br><br>B2) Details of Accidents (Death)<br>*dead at the scene* *died at the scene*<br>*killed in the crash* *who died in the*<br><br>B3) Details of Accidents (Injury)<br>*was taken to hospital* *were taken to hospital*<br>*been taken to hospital* *injured and taken to*<br>*in a serious condition* *in a stable condition*<br><br>B4) Details of Accidents (multifunctional)<br>*in a car crash*<br><br>C. Response to the Accidents<br>*rushed to the scene*<br><br>D. Aftermath of the Accidents<br>*of causing death by* *the road was closed*<br><br>E. The Cause of the Accidents<br>*the cause of the* *was not immediately clear*<br>*an investigation into the* *an investigation has been* |

| F. External Information Source | | |
|---|---|---|
| *officials say at least* | *said in a statement* | *said in a tweet* |
| *the coast guard said* | *told the AFP news* | |
| *told the associated press* | *told Reuters news agency* | |
| *at a news conference* | *from the scene showed* | |

## I. Stance Expressions

For this functional category, the only two bundles identified, *was believed to have* and *what appeared to be*, function as hedges to express a certain level of uncertainty.

In the corpus for this study, the bundle *was believed to have* is either followed by a passive structure "been + past participle" or by noun phrases:

… the accident *was believed to have* been caused by…

…another male, 20, *was believed to have* minor injuries.

As for the other stance bundle w*hat appeared to be*, it appears after prepositions "in" or "of" and is followed by noun and noun phrases:

…a seat covered in *what appeared to be* blood…

…a picture of *what appeared to be* the wreckage of the truck…

## II. Discourse Organizers

Only one lexical bundle fits this category, *as well as the,* and it functions as a topic elaboration or classification bundle to link noun phrases in a sentence that already contains the conjunction 'and' as in:

….. the switch, wheels, axles and suspension systems, *as well as the* track geometry and condition….

## III. Referential Expressions

Referential bundles can be further classified into three sub-categories: identification/focus, specification of attributes and time/place reference.

### *A. Identification/Focus*

There are six lexical bundles that belong to this category:

| | | |
|---|---|---|
| *one of the worst* | *one of the deadliest* | *one of the most* |
| *is one of the* | *were among the dead* | *children were among the* |

As can be seen above, there are lexical bundles that contain the phrase "one of" and those with the word "among".

In the first group, lexical bundles with "one of" also contain superlative forms (worst, deadliest, and most). The bundles *one of the worst* and

*one of the deadliest* are usually preceded by the preposition "in" and they themselves precede the nouns and noun phrases with the head noun "accident" or other words that refers to certain types of accidents:

> …in *one of the worst* accidents to befall….
> …in *one of the deadliest* aviation accidents…

The lexical bundle *one of the most*, in contrast, does not reference the accidents, but rather it occurs in the pattern with a pronoun, a proper noun, or a noun phrase as the subject, followed by the verb *be*, in front of the bundle. After the bundle is information about the antecedent subject:

> It is *one of the most* popular destinations…
> He was *one of the most* revered athletes in the world…
> Greece is *one of the most* popular routes…
> Travel by boat is *one of the most* commonly used methods…

As for *is one of the*, although the lexical bundle itself does not contain any superlative expressions, it is followed by superlative adjectives:

> It *is one of the* worst accidents of its kind in Mexico.
> Egypt's railway network *is one of the* oldest in Africa and the Middle East…

In the second group are the lexical bundles with the preposition "among": *were among the dead* and *children were among the*. For *were among the dead*, the subjects of this verb phrase are references to people without specific names:

> A six-year-old boy and a three-year-old girl *were among the dead*, police said.
> Taiwan's transport ministry said two US citizens *were among the dead*,...

For the other bundle, *children were among the*, what comes after this expression is either the adjective "dead" or it can be noun phrases with specific numbers of people who were killed in the accidents. In this case, what comes before is also specific numbers of children.

> She also said *children were among the* dead.
> At least two *children were among the* 32 killed, and 20 others were injured in the accident…
> …12 *children were among the* some 45 people who died in a fiery bus crash…

### *B. Specification of Attributes*

This functional category is further split into two sub-categories – quantity specification and tangible framing attributes – each of which will be discussed in detail below.

*B1) Quantity Specification*

| | |
|---|---|
| *at least three people* | *more than 100 people* |
| *of the passengers were* | *accidents are common in* |
| *are not uncommon in* | |

Five lexical bundles belong to this category, and they can be divided into two groups: three bundles that quantify people (*at least three people, more than 100* people, and *of the passengers were)* and two that refer to the number accidents (*accidents are common in* and *are not uncommon in*).

For the lexical bundles that quantify people, *at least three people* and *more than 100 people* are on the opposite ends of referencing the numbers of people from the minimum "at least" to maximum "more than." The bundle *at least three people* is used in the context of death in accidents and can occur as objects of a verb:

> …after a migrant smuggling vessel sank, leaving *at least three people* dead and dozens reported missing.
>
> A train crash…has killed *at least three people* and injured dozens more.

Alternatively, it can also be a subject of a clause followed by verb phrases of death or injuries:

> …rescuers at the scene saying *at least three people* were dead.
>
> *At least three people* have died and more than 40 others suffered injuries…

As for *more than 100 people*, it is used primarily as a subject of a clause, in both active and passive voice constructions:

> …*more than 100 people* died in Kebbi state in a similar accident.
>
> …*more than 100 people* were killed when a petrol tanker overturned.
>
> *More than 100 people* were injured in that crash.

The quantity bundle *of the passengers were* primarily comes after quantifiers like "most" or "several":

> Most *of the passengers were* returning home to ….
>
> He said several *of the passengers were* ejected from the vehicle upon collision.

It can also be preceded by a specific number:

> ….three *of the passengers were* guests and two were guides.

As can be seen from the examples above, the verb "were" can be either an auxiliary verb for a continuous aspect or a passive structure or a copula verb followed by a noun or a noun phrase.

The last two lexical bundles in this category – *accidents are common in* and *are not uncommon in* – quantify the frequency occurrences of accidents.

They are followed by proper names of countries or a noun phrase that reference a specific country.

For *accidents are common in*, there are modifiers before the bundle itself to specify what kind of accidents, such as "boat" and "road":

> Boat *accidents are common in* Nigeria,...
>
> Road *accidents are common in* the Andean country…

Similarly, for *are not uncommon in*, what comes before are specific types of accidents as well, such as "fuel truck accidents" or "bus plunges":

> Deadly fuel truck accidents along perilous roads *are not uncommon in* Kenya …
>
> … although bus plunges *are not uncommon in* Peru, especially in the Andes.

These two bundles usually occur towards the end of a news article to provide additional information about the region where the accidents happened.

*B2) Tangible Framing Attributes*

Four lexical bundles fall into theis category and they are in the structure of *the + singular noun + of [a/the]:*

*the body of a* *the driver of the* *the scene of the* *the site of the*

The lexical bundle *the body of a* is followed by the noun "man" and is primarily used in the context of the verbs "spot" or "find" in a passive voice structure:

> ….that *the body of a* man was spotted near the bridge.
>
> *The body of a* man … was found on the river bank…

As for the lexical bundle *the driver of the*, what comes after the bundle are either a type of vehicle (e.g. lorry or car) or a specific description of the vehicle (e.g. Isuzu pickup or eight-carriage train). The phrases with this lexical bundle are primarily subject of a clause:

> *The driver of the* lorry was uninjured…
>
> …*the driver of the* test train had taken the wrong route…

They are also noun phrases functioning as modifiers of proper person names, either in a relative clause or as an appositive:

> Mrs Kelly, who was *the driver of the* car, died at the scene.
>
> Pai Amkrathok, 75, *the driver of the* farm truck, was killed,...

The last two lexical bundles that perform the tangible framing attribute function are *the scene of the* and *the site of the*. Both are similar in that they precede nouns like crash and accident. However, for *the scene of the*, it usually occurs after the preposition "at" and in half of its occurrences, this bundle is used in the context of death or injuries.

> The injured were given first aid at *the scene of the* accident …

> Police said the men, who were all in their 20s, died at *the scene of the* crash.

On the other hand, the bundle *the site of the* is usually used with verbs of movement and after the preposition "to":

> ….senior military figures were travelling to *the site of the* crash.
> …. head to *the site of the* crash with the ministers of health and social solidarity…

### *C. Time/Place Reference*

This is the last subcategory for the function of referential expression. For the lexical bundles in this time/place reference function, they can be divided further into three types: place, time and multifunctional references.

*C1) Place Reference*

The place reference function comprises 11 lexical bundles, which is not surprising considering that locations are important when reporting on accident news.

| | | |
|---|---|---|
| *near the town of* | *near the city of* | *near the village of* |
| *south of the capital* | *off the coast of* | *on a highway in* |
| *a car crash in* | *at the junction of* | *in the middle of* |
| *in the opposite direction* | *were on board the* | |

Location-wise, the 11 place reference bundles can be further divided into three different groups with different usage.

The first group consists of 4 lexical bundles that introduce proper names of towns, cities, villages and countries. Three of these are *near the town of, near the city of* and *near the village of*, which are followed by the names of places that most might not be familiar with, hence the necessity to include the information that those are towns, cities or villages:

> …*near the town of* Kanosh…
> …*near the city of* Petropavlovsk-Kamchatsky.
> …*near the village of* Cekmece.

The other lexical bundle that introduces a city name is *south of the capital*, which is followed by the proper names of the capital cities. It is also preceded by another geographical location with distances in kilometers and miles:

> … into the small village of Cambesse, 180km (120 miles) *south of the capital* Bissau.
> …near the southern rural town of Jhalokati, 250km (155 miles) *south of the capital*, Dhaka.

The second group of place lexical bundles comprise 3 phrases that are followed by a country or a city: *off the coast of, on a highway in* and *a car crash in*.

… crash into the sea *off the coast of* San Diego…
… *on a highway in* northern Japan…
… *a car crash in* Doha…

The final group is four bundles that refer to smaller locations compared to the other two groups: *at the junction of*, *in the middle of*, *in the opposite direction* and *were on board the*.

The bundle at *at the junction of* primarily occurs in the pattern where it is preceded by mentions of an accident and followed by specific road names:

…the collision, which happened *at the junction of* Gloucester Road and Zetland Road…
…road traffic collision *at the junction of* Mayals Road and Fairwood Road.

Similarly, *in the middle of* is used with roads and rivers where the accidents happened:

…it had stopped *in the middle of* the road.
The crash occurred about 11am *in the middle of* the river…

As for *in the opposite direction*, it is used with vehicles and verbs of movements such as:

…a red Ford Fiesta travelling *in the opposite direction* …
The empty train … was going *in the opposite direction* on the same track as the full one.

Even when there is no verb of movement, the noun used still refers to movement:

…in the last hours of a similar-length journey *in the opposite direction* from Rawalpindi….

The last place lexical bundle *were on board the* is used to precede boats, planes or trains and are preceded by the number of groups of people as the subject of this structure:

…. 165 other people who *were on board the* boats heading to…
Ninety-two people, most of them recent army graduates, *were on board the* C-130 Hercules transport aircraft…

*C2) Time Reference*

The two lexical bundles that function as time reference are *at the time of* and *in the early hours*. The bundle *at the time of* is solely used to refer to when the accidents happened and as such, is followed by the nouns referring to the accidents:

…who were on the train *at the time of* the collision.

…suggesting many passengers were standing *at the time of* the crash.

The other time bundle *in the early hours* also refers to the time of the accidents. It occurs in the pattern where the bundle is followed by the preposition "of" and a day of the week:

…the boat sailed from the city of Zawiya *in the early hours* of Tuesday…

…the truck ploughed through several houses *in the early hours* of Sunday,…

*C3) Multifunctional Reference*

Only one bundle, *at the end of*, can perform the dual functions of place and time references. When used as a place reference, it is followed by words related to ways of transportation such as a runway for a plane crash accident:

… that the aircraft crashed *at the end of* the runway.

As a time reference, this lexical bundle preceded varying periods of time, from days to a specific year:

… came *at the end of* a holiday weekend

…. who left F1 *at the end of* 2018, …

## IV. News-Specific Functions

This final functional category has been added to include lexical bundles that are specific to the genre of accident news. There are 6 sub-categories that correspond to the different parts of or perform different functions in a news article.

### *A. Summary of Accidents*

Lexical bundles that function as a summary of accidents are either a general summary or a summary referring to death. They can occur at the beginning of the news in the first paragraph to give a summary of the news being reported or appear towards the end of the news giving a summary of other related news, such as previous similar accidents.

*A1) Summary of Accidents (General)*

Three lexical bundles are part of a general summary of the accidents: *in a collision with, was involved in a* and *been involved in a*.

The bundle *in a collision with* occurs in the same context where death or injury is reported in the news and the bundle itself outlines who or what are part of the accidents:

> A motorcyclist has died after he was involved *in a collision with* two cars.
>
> …died from injuries sustained when the bike he was riding was involved *in a collision with* a car…

The second bundle *was involved in a* overlaps with the bundle *in a collision with* above in seven cases:

> A motorcyclist has died after he *was involved in a* collision with a car.
>
> …died from injuries sustained when the bike he was riding *was involved in a* collision with a car…

Other occurrences of *was involved in a* are summary of other news:

> …its team on Friday night *was involved in a* search for a boat in distress…
>
> …the suspect *was involved in a* domestic disturbance just prior to the incident.

The last general summary lexical bundle *been involved in a* is used in a pattern where it is followed by a mention of one accident when it is a summary of the news being reported:

> Fernando Alonso had *been involved in a* road accident while cycling in Switzerland…

In addition, it is followed by a mention of multiple accidents (e.g. a number of accidents) when it is a summary of other related news:

> They have *been involved in a* number of accidents in recent years.
>
> An-26 planes, …, have *been involved in a* number of accidents…

*A2) Summary of Accidents (Death)*

Lexical bundles that give a summary of death form the largest functional group at 9 lexical bundles, which are divided further into two categories, one with the verb "kill" and one with the verb "die."

In the first group with "kill," there are 5-lexical bundles:

> *people were killed when* *people were killed in*
> *people were killed and* *people have been killed*
> *that killed at least*

Most of the lexical bundles in this group, with the exception of *that killed at least*, are preceded by some form of quantifiers. For *people were killed when*, the quantifying expressions that go before the lexical bundles are exact numbers, the adjective "several," the phrase "more than 100" and the phrase

with "at least" with exact numbers. These lexical bundles are followed by descriptions of what happened in the accidents:

> In October 2018, 189 *people were killed when* a Lion Air Boeing 737 Max jet slammed into the Java Sea…
>
> Several *people were killed when* a Belarussian-owned cargo plane crash-landed in Siberia
>
> In 2009, more than 100 *people were killed when* a petrol tanker overturned…
>
> In 2016, at least 51 *people were killed when* two commuter trains collided…

The lexical bundle *people were killed in* follows a specific number or "at least" with a specific number. What follows the bundle is either a reference to the accident, a time reference (usually a year), or a location:

> Nine *people were killed in* a railway accident early Friday morning in China's Gansu province…
>
> At least 19 *people were killed in* 2018 when a small aircraft carrying passengers from Juba to the city of Yirol crashed.
>
> At least 91 *people were killed in* the capital of Sierra Leone on Friday

For *people were killed and*, it is preceded by either an exact number or "at least" with an exact number and is followed by a report of the number of injuries and then, a reference to the accidents:

> Five *people were killed and* one injured in a helicopter crash at a glacier near Anchorage…
>
> At least 20 *people were killed and* nearly 200 were injured in March when two trains collided…

The lexical bundle *people have been killed* primarily follows a phrase with "at least" with an exact number and only follows three cases of just exact numbers. What follows this bundle can be either a report of injuries or a reference to a specific accident. It is only used as a summary at the beginning of the news:

> At least 26 *people have been killed* after a speedboat packed with passengers collided with a vessel…
>
> At least 40 *people have been killed* and more than 120 injured…
>
> Eleven *people have been killed* and 98 have been injured in the second deadly rail accident in Egypt in a month.

The last lexical bundle in the "kill" group, *that killed at least*, is used after a noun with an indefinite article that refers to the accident and followed by the number of the people who were killed and in some cases who were injured:

> … a train accident *that killed at least* 51 people.

> … a train collision *that killed at least* 19 people and injured 185…

For the second group of lexical bundles that give a summary of death in news, they are 4 bundles with the verb "die":

| | |
|---|---|
| *has died in a* | *people have died in* |
| *died in the crash* | *died in a crash* |

The bundle *has died in a* is used after a subject that is a proper noun of a well-known person or a noun phrase, sometimes with a description. It is followed by a reference to the accidents:

> Alaa al-Siddiq, a prominent dissident Emirati rights activist and critic, *has died in a* car crash near London.
>
> A nine-year-old boy *has died in a* hit-and-run crash in West Yorkshire.

In a similar manner, the bundle *people have died in* also precedes a reference to the accidents, with specific numbers of people before it:

> At least three *people have died in* a plane crash…
>
> at least 60 *people have died in* a boat accident on the Niger River…

Although the last two bundles in this group (*died in the crash* and *died in a crash*) only differ by the definite and indefinite articles, their usage in context is different. *Died in the crash* only appears at the end of the sentence since the reference to the accidents is clear:

> The de facto French Embassy in Taipei confirmed that one of its citizens had *died in the crash.*
>
> …all eight passengers and two crew members on board *died in the crash*.

As for *died in a crash*, it is followed by more modifiers to explain more about the accidents that happened:

> Fourteen people *died in a crash* last month when an Azerbaijani military Mi-17 chopper went down…
>
> A motorcyclist has *died in a crash* involving a van in Hampshire.

### *B. Details of Accidents*

The lexical bundles in this category provide further details of what happened in the accidents. As such, they are further divided into four further categories: general details, details of death, details of injuries and multifunctional ones.

*B1) Details of Accidents (General)*

| | | |
|---|---|---|
| *the accident took place* | *was hit by a* | *crashed into the sea* |
| *and burst into flames* | *they were travelling in* | |
| *as they tried to* | *into the rear of* | *off the road and* |

The 8 lexical bundles above give specific details of how the accidents happened, one of which (*the accident took place*) introduces the time or the place of the accidents, as well as a combination of both the time and the place, and it is always at the beginning of a sentence that functions as the start of the detailed explanation of what happened:

> *The accident took place* at about 8:33pm (12:33 GMT) on -- Monday…
> *The accident took place* near the town of Khuzdar…
> *The accident took place* early on Tuesday near a village in Surat…

The other 7 lexical bundles involve action verbs (5 bundles) or prepositional phrases of such verbs (2 bundles). For example, the bundle *was hit by a* contains the subject that is a person or a vehicle with victims of the accidents and it is followed by a vehicle that caused the accidents:

> A 23-year-old man who *was hit by a* car has died.
> The driver of a small motorcycle was killed when it *was hit by a* much bigger bike…

The bundle *crashed into the sea*, naturally, is a verb phrase for a subject that is an aircraft. It is followed by the prepositional phrase with "off" and a specific geographical location:

> …the fighters, each with one pilot on board, *crashed into the sea* off the island's southeastern coast…
> …the plane had *crashed into the sea* off the Kamchatka Peninsula as it was preparing to land.

The bundle *and burst into flames* would follow another verb phrase with a verb in the meaning of to crash:

> …when her vehicle careered across the Phetkasem highway, ran into a power post *and burst into flames* on Monday…
> The vehicle was driving at high speed late Saturday north of Houston when it slammed into a tree *and burst into flames*…

Although not related to the main action of the accident, the two bundles *they were travelling in* and *as they tried to* provide additional information that is also important to the news. For *they were travelling in*, it is a relative clause modifying the noun phrase that is a vehicle involved in the accident that the deceased in the news were in. The lexical bundle is also followed by a main verb in the sense of to crash:

> ….have been killed after a military helicopter *they were travelling in* crashed in the southern state of Tamil Nadu,
>
> Two men were killed and another hurt when the car *they were travelling in* rear-ended a container truck….

For the bundle *as they tried to*, it is part of the subordinate clause that provides information of the destination or the journey with the main clause being about accidents at sea:

> At least 41 people have died after their boat sank off Tunisia's coast *as they tried to* cross the Mediterranean to reach the Italian island of Lampedusa.
>
> … more than 2,500 people have died or gone missing at sea *as they tried to* reach Europe…

The two preposition lexical bundles in this category, *into the rear of* and *off the road and*, follow specific kinds of verbs. The verbs that come before *into the rear of* are those that mean to violently crash into something:

> …the 22-year-old slammed *into the rear of* another vehicle..
>
> …when it rammed *into the rear of* a Honda Accord…

Meanwhile, *off the road and* is fronted by verbs of movement, such as "go," and other verbs that signify quick movements:

> …after a vehicle he was driving went *off the road and* rolled over several times…
>
> …the car suddenly veered *off the road and* hit a tamarind tree …

*B2) Details of Accidents (Death)*

Four lexical bundles belong to this category where the details on the deaths in the accidents are given:

| | |
|---|---|
| *dead at the scene* | *died at the scene* |
| *killed in the crash* | *who died in the* |

Almost all occurrences (26 out of 29 instances) of *dead at the scene* are preceded by the passive construction "was/were pronounced":

> Mr King was pronounced *dead at the scene*...
>
> The man, believed to be in his 50s, was pronounced *dead at the scene*…

The other three instances of *dead at the scene* are preceded by the passivized form of the verbs "confirm" and "found":

> Forty-nine people were confirmed *dead at the scene*…
>
> …the 19-year-old was found *dead at the scene*…

For *died at the scene*, it mostly functions as a complete verb phrase and occurs at the end of the sentence or the clause. In 4 out of 28 cases, it is followed by the prepositional phrase "of the crash":

> Both men travelling in the Renault, aged 26 and 31, *died at the scene.*
>
> …..the men, who were all in their 20s, *died at the scene* of the crash.

The bundle *killed in the crash* either appear as part of a passive structure or used in a reduced relative clause form as a participial phrase after the demonstrative pronoun "those":

> … both drivers had been *killed in the crash* so …
>
> The identities of those *killed in the crash* were not immediately released.

The last lexical bundle that provides details about death in the accidents is *who died in the*, which forms a relative clause. It occurs primarily after the pronoun "those" and before nouns that reference the accidents, such as "incident" or "crash"

> …almost all of those *who died in the* crash were ethnic Albanians.
>
> …the bodies of those *who died in the* incident were…

*B3) Details of Accidents (Injury)*

The six lexical bundles in this category give details of injuries in accidents:

| | |
|---|---|
| *was taken to hospital* | *were taken to hospital* |
| *been taken to hospital* | *injured and taken to* |
| *in a serious condition* | *in a stable condition* |

Although *was taken to hospital*, *were taken to hospital* and *been taken to hospital* are very similar and only differ from one another by one word, the verb to be, the pattern of occurrences for these three lexical bundles varies slightly and merits separate discussions.

For *was taken to hospital*, the clause with this lexical bundle as the verb phrase usually follows another independent clause where death is reported, usually with the expression "was pronounced dead" and it is followed by the preposition "with" and the type of injury or condition:

> The driver was pronounced dead at the scene and a passenger from the same car *was taken to hospital* with serious injuries.
>
> Mr King was pronounced dead at the scene and Small *was taken to hospital* with a collapsed lung, broken arm, and two fractured ribs.

The bundle *were taken to hospital* primarily occurs in a separate sentence from reports of death, although it is also followed by "with" and details of injuries:

Seven people escaped from the bus and *were taken to hospital* with burns.

Two of the children *were taken to hospital* with minor injuries after the two-vehicle collision,...

As evident from the past participle "been" in the lexical bundle, *been taken to hospital* occurs after the auxiliary "have" to form either present or past perfect. It is followed by a clause or a prepositional phrase beginning with "after" that explains what happened in the accident, sometimes with a prepositional phrase "with" and details of injuries:

Two people have *been taken to hospital* after an open-top bus crashed into a tree…

She had *been taken to hospital* after the crash…

An e-scooter rider has *been taken to hospital* with serious injuries after a collision with a car.

For *injured and taken to*, in some cases it is preceded by the adverb "seriously" and what follows this bundle is either the noun "hospital" or a proper name of a hospital:

The driver of the truck was *injured and taken* to hospital,…

The driver was seriously *injured and taken to* the Queen Elizabeth Hospital.

The last two lexical bundles in this injury detail category, *in a serious condition* and *in a stable condition*, although different in meaning, share similar occurrences where they are preceded by the copula "be" or the linking verb "remain":

Six people were *in a serious condition*…

She was *in a stable condition*…

*B4) Details of Accidents (multifunctional)*

Only one lexical bundle, *in a car crash,* is multifunctional when providing details of either death, injury or both:

The 24-year-old athlete died *in a car crash* …

A young beauty pageant queen has died from injuries received *in a car crash*..

A teenage boy has been killed and another critically injured *in a car crash*…

### *C. Response to the Accidents*

Only one lexical bundle functions as indicating the immediate response to the accidents, which is *rushed to the scene*, and is used in either the active or the passive structures. In both structures, the grammatical subjects are references to firefighters, the police, rescue workers, ambulances or officials:

> Thirty-six ambulances *rushed to the scene*….
> A rescue unit from Non Sung Hospital and police were *rushed to the scene*.

### *D. Aftermath of the Accidents*

Two lexical bundles are used when the aftermath of the accidents are covered in the news. The bundle *the road was closed* is the immediate aftermath and it is followed by a period of time:

> … and *the road was closed* for about eight hours.
> *The road was closed* for some time while investigation work was carried out.

The other bundle, *of causing death by*, is about the aftermath that is far removed from the accident itself. It occurs in the context of an arrest or an official charge in court and are used after such phrases as "on suspicion" or "on charges," followed by the charges themselves that are related to the driving (e.g. dangerous driving or negligence):

> …arrested on suspicion *of causing death by* dangerous driving.
> …to detain the manager on charges *of causing death by* negligence,...

### *E. The Cause of the Accidents*

The four bundles that are related to the cause of the accidents are either a direct reference to the cause or to the investigation:

| | |
|---|---|
| *the cause of the* | *was not immediately clear* |
| *an investigation into the* | *an investigation has been* |

For *the cause of the*, given the nature of the corpus content, it is no surprise that this lexical bundle is followed by the nouns "crash" and "accident." Other nouns that follow this lexical bundle are types of accidents such as "collision" for car and train accidents or "capsize" for boat accidents. It occurs in the context where the cause was not known but officials were in the process of trying to find out:

> ….*the cause of the* crash was still under investigation….

> ….railroad signal specialists will look into *the cause of the* accident…

The bundle *was not immediately clear* falls into two patterns of occurrence. The first pattern is right after the pronoun "it" at the beginning on the sentence and is followed by what is not clear and, in some cases, references to the cause:

> It *was not immediately clear* how many passengers were on board and what caused the derailment.
>
> It *was not immediately clear* if the vessel was the one involved in Tuesday's incident.

The other pattern is at the end of a clause or a sentence, with the phrase "the cause" as the subject of this verb phrase:

> The cause of the crash *was not immediately clear*.
>
> The cause *was not immediately clear*, but President Abdul Fattah al-Sisi ordered an investigation.

For *an investigation into the* and *an investigation has been*, they are both used in the context where an investigation has begun. When *an investigation into the* is used, it appears in a pattern where an authority has launched or begun an investigation structure. It is followed by the noun "cause" or references to the accident:

> Italy has launched *an investigation into the* cause of Sunday's cable car accident…
>
> The Czech Republic's railway inspectorate has launched *an investigation into the* crash.

In contrast to the aforementioned active voice structure, *an investigation has been* is in a passive structure, usually followed by a verb that means to begin, such as "launch" or "open." When appearing with "launch," it is also followed by "into" and the accidents:

> *An investigation has been* launched into the crash…
>
> *An investigation has been* opened for violation of air safety rules,…

### *F. External Information Source*

In this final category, there are nine lexical bundles that indicate information from external sources:

| | |
|---|---|
| *officials say at least* | *said in a statement* |
| *said in a tweet* | *the coast guard said* |
| *told the AFP news* | *told the associated press* |
| *told Reuters news agency* | *at a news conference* |
| *from the scene showed* | |

From the nine above, the first seven lexical bundles include the reporting verbs "say" and "tell" in past simple tense, with the exception of one bundle with the verb "say" in the present simple tense. Except for the bundle *officials say at least,* the pattern of use in news articles of the reporting verb part that contain these bundles are similar: they either precede the reported speech or direct quote part or they can appear at the end following a comma. Here are some examples of the lexical bundles preceding the reporting speech part:

> The Federal Aviation Administration *said in a statement* that it would investigate Tuesday's explosion…
>
> The Fort Worth Police Department *said in a tweet* that anyone involved in a minor accident with no injuries should…
>
> ...the government administrator for Bhasan Char, *told the AFP news* agency that the missing included women…
>
> A spokesman for Border Patrol, Macario Mora, *told the Associated Press* that the immigration status …

The examples below are when the phrases containing the external information lexical bundles follow the quotes or reported speech after a comma:

> …confirmed the death of 10 Mexicans so far," Roberto Velasco Alvarez, the foreign ministry's director for North America, *said in a tweet* in Spanish.
>
> Ninety people stranded on the islet were rescued, *the coast guard said*.
>
> "....we can't even save the ones who are alive," a man at the scene *told Reuters news agency* on Friday.

For *officials say at least*, since it straddles the reporting verb part and the reported speech, it only occurs in the initial position and usually contain modifiers that are the names of a department or a unit (e.g. Railways):

> Railways *officials say at least* 63 people died …
>
> Fire and rescue *officials say at least* four people were seriously injured …

The other two lexical bundles that do not contain the reporting verb are *at a news conference* and *from the scene showed*. The bundle *at a news conference* is used immediately after reporting verbs like "say" or "add":

> Two planes carrying a total of 52 doctors, mostly surgeons, were sent to Sohag, she added *at a news conference* in the province,
>
> "...I just hope he's alright," Justin Thomas said *at a news conference*.

As for *from the scene showed*, it follows references to photos or videos and precedes detailed descriptions of the photos or videos:

Images *from the scene showed* carriages in the tunnel ripped -- apart by the impact,...
… and videos on local media *from the scene showed* wagons with passengers trapped inside and surrounded by debris.

## Conclusion

This analysis of lexical bundles in vehicular news articles categorizes the frequent lexical bundles in the genre of news according to their functions. By creating a new functional category just for lexical bundles in news and filtering the results further into different parts or functions in the news, this allows the list to be of benefit for students who are learning to write in this genre. In addition, frequent context or patterns of usage of each lexical bundle are also identified, which can make this a valuable resource for students in writing. However, as Simpson-Vlach and Ellis (2010) state, such a list should be a "starting point" (p. 502) rather than the materials for classroom teaching or self-study. The subsequent phase from this study would be to design materials for students based on this news lexical bundle list or to allow students to interact with the news corpora directly to maximize their learning. Future research could also focus on the effects of implementing this list or materials developed from it in real classroom settings to investigate the effect of the list or the materials and on how they can help students to better translate the news of this genre from Thai to English.

## About the Author

**Mathurin Leelasetakul**: A lecturer at the Department of English, Faculty of Arts, Chulalongkorn University and a researcher in the Corpus Linguistics for Digital Humanities Research Unit, Faculty of Arts, Chulalongkorn University. Her interests include English language teaching and corpus linguistics.

## References

Ackermann, K., & Chen, Y. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*(4), 235-247.

AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes, 17,* 51–62.

Anthony, L. (2024). AntConc (Version 4.3.0) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software/AntConc

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard and S. Oksetjell (Eds.), *Out of corpora: Studies in honor of Stig Iohansson* (pp. 181-190). Rodopi.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*, 371-405.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*, 263-286.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 12*(2), 30-49.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*, 397-423.

Coxhead, A. (2000). A new Academic Word List. *TESOL Quarterly, 34*(2), 213-238.

Coxhead, A., & Hirsh, D. (2007). A pilot science-specific word list. *Revue Française de Linguistique Appliquée*, *Vol. XII*(2), 65–78.

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

Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional settings. In U. Connor and T. Upton (Eds), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 11–33). John Benjamins.

Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4-21.

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*, 41-62.

Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, *91*(3), 433–445.

Koester, A. (2010). Building small specialised corpora. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 66 - 79). Routledge.

Laosrirattanachai, P., & Laosrirattanachai, P. (2024). Hope you guys enjoyed this vlog: Lexical bundles and moves within English travel Vlog discourse. *LEARN Journal: Language Education and Acquisition Research Network*, *17*(2), 249–278.

Moon, R. 1997. Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy M. (Eds.), *Vocabulary: description, acquisition and pedagogy* (pp. 40-63). Cambridge University Press.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press.

Oxford Learner's Dictionary. (2024, August). *The Oxford 3000 and the Oxford 5000*. https://www.oxfordlearnersdictionaries.com/about/wordlists/oxford3000-5000

Peters, E. & Pauwels, P. (2015). Learning academic formulaic sequences. *Journal of English for Academic Purposes, 20*, 28–39

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use*. (pp. 1-22). John Benjamins.

Simpson-Vlach, R., & Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512.

Siricharoen, A., & Wijitsopon, R. (2020). A corpus-based comparative study of lexical bundles in authentic and textbook English business emails. *LEARN Journal: Language Education and Acquisition Research Network*, *13*(2), 41–63.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

## Appendix A

List of Lexical Bundles

| | Lexical bundles | Frequency |
|---|---|---|
| 1 | *said in a statement* | 44 |
| 2 | *the cause of the* | 42 |
| 3 | *dead at the scene* | 29 |
| 4 | *died at the scene* | 27 |
| 5 | *off the coast of* | 24 |
| 6 | *the driver of the* | 17 |
| 7 | *in a collision with* | 17 |
| 8 | *people were killed when* | 17 |
| 9 | *people were killed and* | 16 |
| 10 | *was taken to hospital* | 16 |
| 11 | *the scene of the* | 14 |
| 12 | *rushed to the scene* | 13 |

| | | |
|---|---|---|
| 13 | *near the town of* | 12 |
| 14 | *has died in a* | 12 |
| 15 | *were taken to hospital* | 12 |
| 16 | *more than 100 people* | 11 |
| 17 | *the site of the* | 11 |
| 18 | *at the time of* | 11 |
| 19 | *died in a crash* | 11 |
| 20 | *people have been killed* | 11 |
| 21 | *of causing death by* | 11 |
| 22 | *told Reuters news agency* | 11 |
| 23 | *accidents are common in* | 10 |
| 24 | *is one of the* | 9 |
| 25 | *south of the capital* | 9 |
| 26 | *was involved in a* | 9 |
| 27 | *been taken to hospital* | 9 |
| 28 | *said in a tweet* | 9 |
| 29 | *the coast guard said* | 9 |
| 30 | *were among the dead* | 8 |
| 31 | *on a highway in* | 8 |
| 32 | *the accident took place* | 8 |
| 33 | *killed in the crash* | 8 |
| 34 | *an investigation into the* | 8 |
| 35 | *was not immediately clear* | 8 |
| 36 | *from the scene showed* | 8 |
| 37 | *one of the deadliest* | 7 |
| 38 | *one of the most* | 7 |
| 39 | *at least three people* | 7 |
| 40 | *a car crash in* | 7 |
| 41 | *at the junction of* | 7 |
| 42 | *near the city of* | 7 |
| 43 | *at the end of* | 7 |
| 44 | *people were killed in* | 7 |
| 45 | *that killed at least* | 7 |
| 46 | *into the rear of* | 7 |
| 47 | *as they tried to* | 7 |
| 48 | *in a car crash* | 7 |

| | | |
|---|---|---|
| 49 | *the road was closed* | 7 |
| 50 | *an investigation has been* | 7 |
| 51 | *officials say at least* | 7 |
| 52 | *told the afp news* | 7 |
| 53 | *was believed to have* | 6 |
| 54 | *what appeared to be* | 6 |
| 55 | *as well as the* | 6 |
| 56 | *children were among the* | 6 |
| 57 | *one of the worst* | 6 |
| 58 | *are not uncommon in* | 6 |
| 59 | *of the passengers were* | 6 |
| 60 | *the body of a* | 6 |
| 61 | *were on board the* | 6 |
| 62 | *in the middle of* | 6 |
| 63 | *in the opposite direction* | 6 |
| 64 | *near the village of* | 6 |
| 65 | *in the early hours* | 6 |
| 66 | *been involved in a* | 6 |
| 67 | *people have died in* | 6 |
| 68 | *died in the crash* | 6 |
| 69 | *crashed into the sea* | 6 |
| 70 | *and burst into flames* | 6 |
| 71 | *off the road and* | 6 |
| 72 | *they were travelling in* | 6 |
| 73 | *was hit by a* | 6 |
| 74 | *who died in the* | 6 |
| 75 | *in a serious condition* | 6 |
| 76 | *in a stable condition* | 6 |
| 77 | *injured and taken to* | 6 |
| 78 | *at a news conference* | 6 |
| 79 | *told the associated press* | 6 |