"Devious Fungus" and "Genetically Engineered Chestnut"

A Corpus-Assisted Ecolinguistic Analysis of the US-American Biodiversity Discourse

Janine Aloe (Stockholm University/University of Tartu)

Abstract *This chapter analyzes US-American mass media discourse on biodiversity loss with the help of ecolinguistics and critical discourse analysis. It examines how the language in the two ideologically distinct newspapers, The Washington Post and The Wall Street Journal (2022–2024), shapes perceptions of non-human life, particularly underrepresented fungi and plants. Keyword analysis using a specialized corpus on extinction and endangerment shows lexical overlaps as well as contrasts between the two media outlets. Concordance analysis of the fungi and plant-related keywords fungi, milkweed and chestnut suggests a rather instrumental, reductionist portrayal which backgrounds their agency and the interconnectedness of all living beings.*

Keywords *biodiversity, corpus linguistics, ecolinguistics, media discourse, species extinction*

Schlagworte *Artensterben, Biodiversität, Korpuslinguistik, Mediendiskurs, Ökolinguistik*

1 Introduction

Life on Earth is made up of interdependencies and intricate relationships among species and ecosystems, relationships strained by accelerating ecological crises. With seven of nine planetary boundaries now exceeded – including the recently crossed threshold of ocean acidification – the Earth is drifting further from the stable conditions that have long supported the diversity of life (Potsdam Institute for Climate Impact Research 2025). More than three-quarters of the planet's vital functions are no longer within a safe operating space, raising the risk of irreversible tipping points across oceans, forests, and the atmosphere, according to scientists from the Potsdam Institute for Climate Impact Research (PIK).

From the Club of Rome's early warning in *The Limits to Growth* (1972) to the United Nations' Agenda 2030 and its Sustainable Development Goals (UN 2015), scientists and policymakers have repeatedly highlighted the interlinked crises of biodiversity loss, cli-

mate change, and the transgression of planetary boundaries. But these challenges are not only scientific or political; they are profoundly discursive. Media language shapes public awareness and understanding of ecological crises such as species loss, frames what counts as urgent action, and influences whether non-human life is seen as vibrant and agentic or as passive resources for human use. In the Anthropocene, when humans' ecologically destructive behavior threatens "life on land" and "life below water" (UN 2015), scrutinizing the linguistic representation of non-human nature in mass media is therefore not a peripheral concern but a crucial step toward fostering ecological responsibility and motivating transformative change.

This article uses corpus linguistic and ecolinguistic methods to examine patterns of representation of biodiversity (loss), firstly through keyword analysis, followed by a qualitative, more in-depth analysis of selected keywords in a specialized corpus of US-American newspaper articles covering the (impending) loss of species or ecosystems published between 2022 and 2024. The initial keyword analysis, conducted with the corpus analysis tool Sketch Engine, reveals the 'aboutness' of a corpus, whereas a concordance analysis of selected keywords – in this case, *fungus*, *milkweed*, and *tree* – shows differences in their portrayal, possibly laying bare speciesist tendencies in media coverage. This corpus-linguistic approach to extinction and endangerment of non-human species deals with the question whether the specific language choices applied in news reporting promote a respectful treatment of non-human species or support an ideology that embraces an instrumental view of non-human nature and ultimately, promotes anthropocentric, ecologically harmful attitudes and behaviors. A better understanding of such linguistic processes helps to build a better understanding of why major ecological crises, such as the crossing of planetary boundaries, are often ignored in today's public discourse.

Against this backdrop, the present chapter pursues these specific aims: Firstly, the objective of this study is to identify and analyze lexical salience in media discourse surrounding species extinction and lexical differences and similarities between two media outlets through keyword analysis. Secondly, my study attempts to investigate how media representations differ depending on the species or groups of organisms, specifically focusing on plants and fungi.

2 Ecolinguistics

Broadly speaking, ecolinguistics interrogates how language shapes how we think about other species and the environment (Stibbe 2021). Einar Haugen originally defined language ecology as "the study between any given language and its environment" (Haugen [1972] 2001, 57). While Haugen ([1972] 2001, 57) used ecology as a metaphor and understood the environment as the "the society that uses [a language] as one of its codes", there has been a shift towards more discourse-focused strands that analyze the language of ecology. Recently, ecolinguistics has come to be viewed as not merely the analysis of the effect of language on our relationship with the environment, but also the investigation of language as a part of ecosystems (Stibbe 2021, vii). Ecolinguistics emphasizes that language "function[s] to reproduce, normalize, and perpetuate a range of ideologies, practices, beliefs, and attitudes that are harmful to ecological well-being" (Poole 2022, xi) and

it aims to detect ecologically destructive language practices. In parallel, ecolinguists attempt to find and promote what Arran Stibbe calls *beneficial discourses* or stories that convey an ecocentric outlook which sees humans as part of the natural world and attributes intrinsic value to non-human species (Stibbe 2017, 502–503). *Destructive discourses*, on the other hand, produce an "unecological fragmentation" between humans and the environment (Fill and Mühlhäusler 2001, 6). For example, the use of terms like "non-human animal" challenges the human-animal dichotomy that separates humans from all other animals.

Ecolinguistics is considered as either a part or an extension of critical ciscourse analysis (CDA). Ecolinguistics shares CDA's goals of creating awareness and ultimately challenging language forms that perpetuate inequality and exploitation. While CDA traditionally focuses on the discrimination or marginalization of disadvantaged groups, ecolinguistics attempts to raise awareness of linguistic features that subordinate or erase non-human species and ecosystems and thus extends the concepts of justice to the well-being of all living beings (Poole 2022, 3). Douglas Mark Ponton (2024, 8), therefore, argues that ecolinguistics goes beyond CDA by reflecting upon language as an ecosystem as well as interrogating and underlining how language plays a role in humans' life-sustaining relationships with other living beings and the natural world (Stibbe 2021, 84).

3 Analyzing Non-human Nature in Discourse: Corpus-Assisted Ecolinguistics

Within critical discourse studies quantitative approaches are increasingly popular. Andrew Goatly (2002) and Robert Poole (2022) provide good examples of corpus linguistic approaches applied to the interrogation of discourse through an ecocritical lens. Corpus-assisted methods can help lay bare linguistic patterns that reflect how humans think about non-humans and the environment and explain how such language use reproduces attitudes and beliefs that contribute to violent practices towards non-humans.

Goatly (2002) employs critical discourse analysis (CDA) and systemic functional grammar (SFG) to investigate how the BBC World Service radio broadcasts represent nature, focusing on lexical and syntactic patterns that convey agency of ten classes of natural 'objects' such as insects, birds, plants and disease. The analyzed data consists of a corpus of 2.5 million words gathered in 1990–1. He bases his analysis on the Whorfian hypothesis that "the language we speak predisposes us to perceive, think (and act) in certain ways" (Goatley 2002, 2). The study finds that nature is predominantly depicted as passive, domesticated and exploited. Only weather and disease are portrayed as agentive when posing a threat to humans. Collocational and grammatical analyses reveal a strong anthropocentric bias and a reproduction of the nature-culture distinction. The research underscores that the BBC's framing of nature aligns with news values that prioritize human perception of threats, thereby shaping public understanding of nature as primarily passive or dangerous. In order to show that other kinds of representation are possible, Goatly analyzes Wordsworth's *The Prelude* where nature is more communicative and active. He concludes by positing that BBC's language justifies marginalizing and exploiting nature while Wordworth's poetry represents nature as powerful and valuable

apart from its usefulness to humans. While the genres and corpora are very different, the key claim is that different linguistic resources exist.

Other scholars, too, have attempted to find inspiration from fiction. For example, Poole's study (2022) analyzes ecological fiction. His corpus-assisted ecolinguistic analysis of Richard Powers's Pulitzer Prize winning novel, *The Overstory* (2018), highlights language that promotes ecological awareness. His keyness analysis that was conducted by comparing linguistic items in his corpus of the novel to a larger fiction reference corpus revealed a significant proportion of lexical items referring to nature and, particularly, trees. To further explore the depiction of trees, a concordance and collocation analysis showed the agency and animacy attributed to trees in Powers's novel, in stark contrast to dominant discursive representations of trees. Verb collocates such as *listens* and *feels* portray trees as sentient, communicative and social – a view of trees that is increasingly acknowledged within the scientific community (e. g. Simard 2018). His study demonstrates that literary works like *The Overstory* can serve as powerful tools for ecological consciousness, challenging the dominant portrayal of nature as passive, inanimate background. The findings suggest that integrating corpus-assisted methods with ecolinguistics can help identify beneficial discourses that apply language that promotes an ecological consciousness of multispecies interconnectivity and the intrinsic value of non-human species. While the analysis of a single work of fiction cannot be compared to a corpus of media texts, the study is important in showing the richness of linguistic resources that could also, potentially, find their ways into media discourse.

The following section explains my methodological approach (combining corpus linguistics with ecolinguistics). I then describe the way in which I collected the data and compiled my corpus. This is followed by two results sections, the first takes a general 'broad-brush' approach that focuses on keywords in my two subcorpora to find out what is lexically salient. The second is a more detailed analysis that concentrates on the concordances of selected keywords representing the two kingdoms, fungi and plantae: *fungus*, *milkweed* and *chestnut* (Baker et al. 2013: 257).

4 Corpus and Method

4.1 Methodology

This study adopts an ecolinguistic corpus-assisted discourse analytical approach, combining the empirical reliability of corpus linguistics (CL) with the critical perspective of ecolinguistics and CDA (Bray 2024, 28). Corpus-assisted discourse studies (CADS) explore language as social practice through the analysis of large, principled collections of authentic texts – *corpora* – to uncover patterns that might otherwise remain obscure in close reading (Baker et al. 2013, 259; Poole 2022, 13; Gillings et al. 2023, 1). CADS helps to identify and analyze patterns throughout discourses that "sustain and reproduce the social status quo" (Fairclough and Wodak 1997, 258) and the ideology these discourses (re)produce (Poole 2022, 18). The use of electronically encoded data together with computer-based analysis offers "a useful methodological synergy" (Baker et al. 2008) and helps counter the charge of 'cherry-picking' data to fit preconceived arguments and ideas

(Baker et al. 2013, 259). Although corpus methods have some researcher bias, they can reduce subjectivity as triangulating quantitative and qualitative techniques provides "more sophisticated forms of analysis" and strengthens the validity of interpretations (Subtirelu and Baker 2018, 109). Therefore, Paul Baker et al. (2008) argue that a combination of quantitative and qualitative approaches is fruitful when analyzing discourse.

Overall, my methodology combines corpus-driven and corpus-based approaches. While the former attempts to approach the data without any preconceived hypotheses and lets the analysis be driven by frequency and salience, in corpus-based approaches the analyst is testing preexisting ideas or theories (Baker et al. 2013, 259; Subtirelu and Baker 2018, 109).

Thus, in the first step of the analysis I extract keywords from both subcorpora to reveal lexical salience, followed by a qualitative analysis of the usage of selected keywords. The selection of specific keywords is based on examples of the kingdoms *Plantea*, *Fungi* and *Animalia* in the corpus. The term keyword describes words which occur statistically more frequently in one corpus than in a second corpus (Subitrelu and Baker 2018, 113; Gillings et al. 2023). The keyword lists can reveal whether some words are unique in one dataset and lexical differences between two corpora. Since keywords can offer insights into which lexical items distinguish a corpus from a reference corpus and can give clues about the "aboutness and style" of a corpus, it functions as a good starting point of corpus analyses (Scott and Tribble 2006, 55; Poole 2022, 97). Keywords are ranked by their keyness score which measures the degree to which a word appears more frequently in one corpus than in the reference corpus, with those at the top of the list displaying the most significant differences in frequency (Subtirelu and Baker 2018, 113; Gillings et al. 2023). Since keywords are derived by statistical means, this method counters a common criticism of critical discourse analysts 'cherry-picking' small parts of the data that fit their predetermined ideas (Widdowson 2004, 102; Baker et al. 2013, 258). As keyword lists can only reveal the differences between two corpora, the researcher may miss what the two corpora have in common. To spot the similarities as well, it can be fruitful to compare two corpora against a third reference corpus (Gillings et al. 2023). Hence, I will compare my two subcorpora separately against the reference corpus, the Corpus of English broadsheet newspapers 1993–2021, which consists of around 2 million articles from various English language newspapers such as *The Guardian*, *New York Times*, *The Sun* and *Times of India*.

The corpus analysis tool Sketch Engine, described by Paul Rayson and Amanda Potts (2021, 134) as "extraordinarily powerful, with part-of-speech tagging and lemmatisation on a huge number of languages", is used in the present analysis to calculate the keywords. Sketch Engine allows users to use available reference corpora such as the above-mentioned English language newspapers corpus (SiBol), as well as uploading own corpora onto an online database.

According to Nicolas Close Subtirelu and Paul Baker (2018), two measures are relevant besides the keyness score. Since my reference corpus has a much larger word count than my target corpus, it is necessary to also include relative frequency which in this case refers to the number of times a word occurs per million words. For example, even though the word *bison* appears in my *Washington Post* subcorpus only 105 times, whereas *bison* occurs 981 times in the reference corpus, due to the significant difference in corpus

size, the relative frequency of the word is still much higher in the WP corpus compared to the reference corpus.

Secondly, it is important to look at the dispersion of a word across different texts in the corpus, which can be examined with Sketch Engine's function average reduced frequency (ARF). ARF is particularly relevant when it comes to small corpora like mine. A word might appear numerous times within a single article, which leads to high keyness score and relative frequency, and this could be misleading concerning the salience of the word within the entire corpus. In case of an even distribution of the token within the corpus, ARF and absolute frequency have similar values.

In the second corpus-based part of the analysis I employ concordance analysis to closely examine the usage of specific keywords to identify language features, patterns and nuances of the contexts they are used in. I will pay attention to framing; the term can be defined as "mental structures that shape how we see the world. As a result, they shape the goals we seek, the plans we make, the way we act, and what counts as good or bad outcome of our actions (Lakoff 2014, xi).

In the following, I present a basic outline of the methodology based on Subtirelu and Baker (2018, 110) and I will elaborate below on the corpus collection while the corpus-based and corpus-driven approaches that I employed are explained in more detail below.

Table 1: Overview of research steps

Step	**Description**
1. Research topic and questions	A general research question was formulated before. The question was refined after steps 2 and 3. The final version: What lexical salience can be detected in both subcorpora? How do the representations of *fungi*, *milkweed*, and *chestnut* differ across the corpus?
2. Corpus collection	Texts were collected, 'cleaned' and divided into two subcorpora based on whether they were published by *The Washington Post* or *The Wall Street Journal*.
3. Quantitative analysis (keyword analysis)	Keywords were extracted from both corpora.
4. Qualitative analysis (concordance analysis)	I closely examined the usage of specific keywords identified in step 3 in context through concordance analysis.

4.2 Data

I manually compiled a corpus of newspaper articles on the topic of biodiversity loss by using the search option of the widely circulated newspapers *The Washington Post* (WP) and *The Wall Street Journal* (WSJ) using the search strings 'extinct', 'endangered', and 'biodiversity'. US-American newspapers were chosen both because of the wide spread of En-

glish language materials outside of the USA and because of the impact US political discourse has on global discourse on the environment. I then selected all the articles that reported on the (impending) loss of (an individual) species that were published in the years 2022–2024. My final corpus consists of 382,142 words and 363 articles in total, of which 219 articles were published on the online version of WP, and 144 articles were retrieved from WSJ.

My corpus consists of two subcorpora which each comprise articles on species extinction and endangerment that were published by two comparable media venues: WSJ and WP are both large, prominent media outlets and offer a good contrast concerning audience demographics, socioeconomic viewpoints and foci. While WSJ has a rather conservative and neoliberal outlook, with a focus on business interests, WP is more liberal and progressive and (used to) pride(s) itself on its vast coverage of climate-related issues (Buzbee 2022; Campbell 2026).

5 From Keywords to Concordances: Framing Extinction

5.1 Keyword Analysis: Semantic Patterns in Species Loss Coverage

To conduct my keyword analysis, I limited the keywords to only those words that appeared at least once per thousand words and in at least 2% of all texts (Subtirelu and Baker 2018, 113). A 2% dispersion threshold is rather low but given the modest size and focused nature of the corpus, I consider a low ARF score as sufficient to yield meaningful results, especially since this study aims to uncover subtle patterns in how biodiversity is represented in the media. To give an example, the word *vaquita* has one of the lowest ARF scores in the keyword list of the WP subcorpus – the term appears 54 times in total in 5 articles within the 3-year time span that my corpus covers. Considering the large number of animal species and 90 species of cetaceans I find this number of occurrences noteworthy. Lastly, I filtered out the keywords *endanger/ed*, *extinct/tion*, *biodiversity* and *specie/s* from the final keyword lists.

I compiled keyword lists for my two subcorpora by identifying the fifty highest ranked words respectively (see Appendix). In the WP corpus, salient items include numerous species names such as *vaquita*, *bison*, *firefly*, *condor*, and *whale*, alongside terms connected to ecology (*ecosystem*, *habitat*, *invasive*), and human actors (*biologist*, *ecologist*, *conservationist*). The list also features two plant-related keywords (*milkweed*, *chestnut*) and a general reference to *fungus*.

The WSJ keyword list, by contrast, contains both species names (*manatee*, *koala*, *bison*, *whale*, *grizzly*, *panda*), and ecological or conservation-related lexis (*ecosystem*, *habitat*, *deforestation*), but is also notable for its inclusion of governance- and economy-related terms (*COP15*, *fossil-fuel*, *nonprofit*, *turbine*, *wind-farm*). Unlike the WP corpus, no plants or fungi appear among the top fifty keywords.

To make the contrast more transparent, I grouped the keywords into semantic categories without using parts of speech tagging, focusing solely on their semantic content (Table 2). Some similarities that were revealed are the occurrence of an iconic extinct animal species – the *ivory-billed woodpecker* – as well as the overrepresentation of mammals

in both subcorpora. Each corpus featured 10 references to mammals, with culturally significant species like *bison*, *grizzly*, *coyote* and *whale* appearing in both lists. This finding is not surprising, as Ursula Heise (2010, 48) posits that public discourse on biodiversity loss tends to revolve around a relatively small group of animals, among these most often the large and impressively beautiful mammals are foregrounded.

Another point of overlap lies in the references to insects in terms of their ecological function, their role as *pollinators*. This framing hints at an inherent anthropocentrism of media coverage, as the examples below suggest that the importance of insects such as bumble bees is emphasized primarily in relation to their utility.

> [1] "Federal wildlife officials announced Wednesday they will consider adding 10 new species to the Endangered Species Act, including a big bumble bee that serves as a key **pollinator** across the United States." (WP)

> [2] "**Pollinators** play a role in up to $577 billion worth of global crops cultivated each year, by the United Nations' Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. At least 70% of all crop species rely on insect **pollination**, Dr. Ryalls said." (WSJ)

The example of the WSJ illustrates the commodification of insects by highlighting their economic value within global agriculture. In line with Stibbe's (2014) theory on erasure – the "suppression, backgrounding, exclusion, abstraction, and, in general, any means by which texts draw attention away from certain participants or areas of life" (Stibbe 2021, 141) –, the animals are being erased in two different ways: by using the metonym *pollinator* and, thus, referring to them merely by their function within ecosystems. Secondly, such economic frames can also be understood as a form of erasure as the species' uniqueness and interdependencies between all living organisms is overlooked and "replaced with a distorted version of themselves (the stock of biological resources)" (Stibbe 2014, 589). In the WSJ example, words such as *crop*, *cultivate* as well as a concrete sum of money conjure up a commercial and economic frame that reproduces the dominant representation of nature as a resource for us to exploit. In the WP example, the economic value is not directly mentioned and a specific species is named. However, the bumble bee was specifically singled out here not for its own sake, but because of its crucial role in the pollination process and its role as a keystone species. Secondly, referring to species as 'pollinators' activates an ecosystem service frame, foregrounding their role as ecosystem service providers not as agents embedded in complex and interdependent multispecies relationships.

Table 2: Keywords divided into semantic categories

	The Washington Post	***The Wall Street Journal***
Extinct animals	*(ivory-billed) woodpecker*	*(ivory-billed) woodpecker*
Mammals	*vaquita* *bison* *pup* *coyote* *mammal* *whale* *black-footed ferret* *rhino* *grizzly* *cheetah*	*manatee* *bison* *koala* *coyote* *mammal* *whale* *humpback* *panda* *grizzly* *marsupial*
Birds	*(ivory-billed) woodpecker* *condor*	*(ivory-billed) woodpecker* *condor* *bald eagle* *migratory (birds)*
Invertebrates / Insects	*firefly* *pollinator* *insect* *pollinate*	*pollination* *pollinator* *migratory (butterflies)*
Amphibians	*amphibian* *salamander* *frog*	
Reptiles	*turtle* *alligator*	
Plants	*milkweed* *chestnut*	
Fungi	*fungus*	
Humans	*biologist* *ecologist* *Salwan* (*Georges*; photojournalist) *conservationist* *veterinarian* *paleontologist*	*lobsterman* *biologist* *birder* *ecologist*

	The Washington Post	***The Wall Street Journal***
Ecological and biological concepts	*ecosystem* *habitat* *wildlife* *invasive* *nonnative* *die-off* *imperil* *hatchling*	*ecosystem* *habitat*
Conservation and biodiversity governance	*conservation* *IUCN* (the International Union for Conservation of Nature)	*Audubon* (society; environmental organization) *conservancy* *COP15*
Scientific terms	*geologic* *fossilize* *deforestation*	*climate change* *greenhouse-gas* *carbon* *fossilize*
Environmental technologies		*wind-farm* *turbine* *traps* *padlock*
Energy and business-related terms		*fossil-fuel* *Nestlé* *non-profit* *nature-related*
Places / landscape	*Maui* *Smithsonian* *Guam* *zoo* *grassland*	*Tasmanian* *Maine* *Rainforest*

The keyword lists also reveal several differences between the two subcorpora. While fish and invertebrates other than insects (of which merely migratory butterflies and fireflies are mentioned) are missing on both lists, the keyword lists diverge in their scope of animal representation. The WP corpus reveals a more balanced portrayal across various animal classes, with keywords referencing amphibians, reptiles, plants and even the word *fungus* appears as a keyword. In contrast, the WSJ corpus limits its focus primarily to (non-human) mammals, birds, humans and insects. What stands out in the WSJ keyword list is the presence of terms that fall under technological or business-related categories, such as *nature-related*, *fossil-fuel* and *wind-farm*, categories which are absent in the WP corpus. These terms again underline the media venue's neoliberal outlook and the commodification of nature with this emphasis on the global economy depending on well-functioning ecosystems.

> [3] "'Nature risk is financial risk,' said Elizabeth Mrema, co-chair of TNFD. Some $44 trillion of global economic value is moderately or highly dependent on nature, according to the World Economic Forum, while the World Bank has warned that the collapse of natural systems could wipe $2.7 trillion a year from the global economy by 2030. 'Businesses today are inadequately accounting for **nature-related** dependencies, impacts, risks and opportunities,' said David Craig, TNFD's other co-chair." (WSJ)

Building on Stibbe's (2012) concept of discursive erasure, in the following I will focus on the absences or underrepresentations, of important aspects that may have been ignored, sidelined or actively excluded from the discussion. Stibbe draws attention to the twofold erasure of species as they are not just vanishing in the physical sense of dying out but are also being erased from our collective consciousness. Their disappearance often happens unnoticed and unchallenged. As Heise (2010, 48) observes, especially "crustaceans, worms, fungi, and bacteria remain excluded from the discussion, as do plants" and dominant discourses surrounding extinction tend to feature only a very limited part of the fauna and flora. This underrepresentation of certain groups of species can even be observed in scientific discourse – the IUCN Red List, the most comprehensive source on global extinction risk, included until 2008 all of the known mammal and bird species, but only around one tenth of known fish species, a small fraction of the insect and fern species, and merely one species of fungi – by 2020 this number has increased to 285 and yet still only 0.2 % of fungi and 10 % of plant species are so far assessed on the Red List (Heise 2016, 24; Lughadha 2020, 391). As fungi and plants both make up their own kingdoms and have immense taxonomic diversity, as well as playing foundational roles in ecosystems, these absences or marginalizations are quite significant. While animal extinctions receive far greater attention, a better understanding of the endangerment of plants and fungi is not only vital to safeguard vulnerable species and habitats, but also to harness plant- and fungal-based solutions in mitigating environmental crises such as climate change (Lughadha 2020, 390). Eimar Nic Lughadha et al. (2020, 400) further emphasize that "[p]lant and fungal biodiversity support all life on earth, meriting careful stewardship in an increasingly uncertain environment".

Therefore, I will undertake a closer, qualitative analysis of the specific keywords from the categories fungi and plants and conduct a more in-depth analysis of the respective concordance lines to examine more subtle patterns and nuances in their portrayal. The two plant species that featured in the list were *milkweed* and the (American) *chestnut* (the latter also appeared on place 68 on the preliminary WSJ keyword list before applying the ARF limit), whereas for fungi only the general umbrella term *fungus* occurred, with no specific species from this diverse kingdom represented.

5.2 Qualitative Concordance Analysis of Fungi and Plants

According to Baker (2006, 71), one of the most effective methods for a more in-depth, qualitative analysis is concordance analysis. Concordances are defined as instances of a search item in its immediate co-text, presented with a defined amount of co-text (the acronym *KWIC* stands for *Key Word in Context*) to the left and right of the search term (Baker at al. 2008, 279). The term *concordance* tends to be understood as the entire list of

occurrences, while a single citation is referred to as *concordance line*. Sketch Engine has a built-in concordance tool which lists the hits in chronological order of their appearance in the corpus and allows to extend the context window beyond the immediate words on the left and right of the word under investigation (Wulff and Baker 2020, 161–163).

Firstly, I examine the representation of *fungus* which appeared 92 times as a noun, as well as 8 times as an adjective (*fungal*) in my corpus, with most occurrences in the WP subcorpus and only 6 occurrences in the WSJ subcorpus. I used Sketch Engine's concordance tool to get a random sample of 50 concordance lines to receive a smaller, and yet representative sample. Sketch Engine presents the randomized lines in the same order as they appear in the corpus.

In the concordance lines, a striking pattern emerges: fungi are predominantly construed as disease, as a *devious* pathogen with the verbal collocates *infect*, *spread*, *attack*, *wipe out* and *devastate*. Modifiers that co-occur with *fungus* are *marauding*, *killer* and *deadly*, and the compound word *fungal infection* also appears four times in my corpus. Most occurrences describe the fungus as a destructive agent that is responsible for the chestnut blight and a disease killing amphibians and bats. The language stresses lethality, spread and global reach. Secondly, the culinary and economic value of e. g. truffles is highlighted in a few references, as well as occasional references to fungi in herbaria or as research resources.

> [4] "[…] other mammals, reptiles, amphibians, worms and arthropods – and all kinds of pathogens, such as viruses, bacteria and fungi. The analysis reinforced the findings of many of those inquiries: that a hotter world of is one that is more hospitable to many parasites, and less so to humans and other life." (WP)

However, there are also several references to *mycorrhiza*, a symbiotic relationship between fungi and plants which Anna Tsing (2012, 143) refers to as "the most intriguing interspecies companionship". While at a first glance, the corpus appears to acknowledge the interdependence of fungi and flora, when taking the dispersion into account, a different image emerges – even though the term appears several times in the corpus, it only features in a single article titled "How a few bags of dirt could help make the planet more resilient to climate change" which is a report about a research team assessing mycorrhizal fungi in the Kazakh steppe. As biologist Merlin Sheldrake (2020, 4) explains, "[o]ver 90 percent of plants depend on mycorrhizal fungi – from the Greek words for fungus (*mykes*) and root (*rhiza*) – which can link trees in shared networks sometimes referred to as the 'Wood Wide Web'". Therefore, mycorrhizal networks are key to understanding how ecosystems function, and yet their essential role in making soil, transferring nutrients to plants, decomposing dead matter, and digesting pollutants, that is, their "myriad agencies" (Finzi 2024, 2), tends to be overlooked in the studied corpus.

The prevailing erasure or marginalization of these complex fungal agencies in news reporting reproduces the dominant discourse that presents fungi "as either harbingers of disease and degradation, or primarily instrumental or aesthetic objects-in-the-world, reinforcing binary notions of anthropic exceptionalism" (Finzi 2024, 1) and perpetuates a larger cultural reluctance to reject the false dichotomy between the human and the nonhuman. Joshua Finzi (2024, 1) sees our ignorance concerning the interdependencies of

living beings as a key driver of the crises now destabilizing the climate, and with it, global biodiversity.

Secondly, I investigated the representation of plants in my specialized media corpus of biodiversity loss. Looking closer at the concordance lines of *milkweed* which appeared 25 times in the WP corpus in eight different articles, two distinct discursive framings emerge. In the majority of occurrences, *milkweed* is mentioned in reference to monarch butterflies as the numbers of this iconic, charismatic butterfly species have dropped dramatically and milkweed is their only food source and the only plant on which they lay their eggs (WWF 2025):

> [5] "Pesticides and herbicides threaten not only the insect itself but also milkweed, which monarch larvae need to live." (WP)

Thus, the focus is on the protection of the butterfly species and milkweed is mostly given a utilitarian value as food source and habitat for the insects. This framing promotes the view of plants as inanimate resources and thus such discourse contributes to what Stibbe (2012) refers to as erasure since the plant's intrinsic value is sidelined in favor of its service to a more symbolically resonant animal.

However, there are also instances that resist this pattern. Specific milkweed species are named, such as *butterfly, swamp* and *prostrate milkweed*. Naming specific species is "likely to be fundamental in the understanding of these non-human identities and in the recognition of what sets them apart from others, therefore functioning as central drivers of value estimation and preference" (Blades 2020, 2). Particularly the following example [6], in which a variety of plants are mentioned, can be considered an instance of beneficial discourse, since several plant species are made visible and their part of a vibrant ecological community is emphasized:

> [6] "Now, years later, it's a mature pollinator meadow, bursting with color and a crazy patchwork of goldenrod, milkweed, sumac, aster, coneflower, false indigo, black-eyed Susan, pokeweed, sunflower, wild bergamot, bee balm, coreopsis, ironweed and mountain mint." (WP)

In the article, the author gives recommendations on how to turn your garden into a biodiverse haven and explains why planting native trees and shrubs and being mindful of bugs will have a positive impact on the whole ecosystem that we co-inhabit. Still, even in these cases, the plants are most often linked back to their role in sustaining pollinators.

Turning to the second plant-related keyword, *chestnut*, the third concordance analysis aims at investigating the representation of arboreal life using the example of this specific iconic American tree species that has, in fact, been functionally extinct for about a decade. Its immense cultural and historical significance is illustrated in example [7]:

> [7] "Billions of American chestnuts once presided over canopies from Maine to Mississippi, drizzling the forest floor with nuts for people and animals to eat and providing straight-grained, rot-resistant wood for building the young American nation." (WP)

While *tree/s* appears 485 times in my corpus, the word *chestnut* has 112 hits, with 69 in the WP and 43 in the WSJ subcorpus. The term *chestnut* has two meanings: it can refer to the tree species or its edible nuts. However, most concordance lines focus on the former, and more specifically, on the subgenus *American chestnut*. As the example above illustrates, the tree species is of high cultural and historical value. Additionally, their ability to produce food for human and nonhuman animals, as well as being a source of timber, adds to their status. This culturally significant tree species is only mentioned in five articles in total, but most of these articles are detailed accounts of efforts to resurrect the tree by genetically modifying it and thus, making it resistant to the fungal infection that once "nearly drove the tree out of existence" (Grandoni 2022). Examining the most frequent words co-occurring with *chestnut* with the help of Sketch Engine's collocation tool Word Sketch, the most frequent modifiers (besides *American*) are *blight-resistant* and *transgenic* which already hints at human technological intervention as the central theme. To take a closer look at the nuances of the portrayal of this once abundant tree species in my corpus and to examine whether it is presented as an active living being or merely static background, the main focus of my concordance analysis lies on the verbs. This approach is informed by Michael Halliday's (1990) insights into the representation of agency by the English grammar. Halliday contends that agency is attributed to entities along a continuum from most to least agentive with humans being granted most agency (Poole 2022, 102). Halliday ([1990] 2001, 193) posits that grammatical structures "conspire […] to construe reality in a certain way; and it is a way that is no longer good for our health as a species" typically representing humans as the most animate of beings, while forests and trees, despite their vital ecological functions, are construed as passive background entities. He emphasizes this erasure of agency, noting that although forests actively harbor animals, prevent flooding, and provide oxygen, they are rarely represented as active participants.

> [8] "Gene editing could **revive** a nearly lost tree. Not everyone is on board. **Saving** the American chestnut could restore a piece of history, resurrect a lost ecosystem and combat climate change." (WP)
>
> [9] "His team would also need help to **spread** blight-resistant chestnuts across hundreds of miles of mountains." (WP)
>
> [10] "For nearly a decade, Jared Westbrook has worked on resurrecting the American chestnut, an iconic tree that nearly **vanished** from the United States a century ago." (WP)
>
> [11] "A Chinese chestnut tree is crossed with pollen from an American chestnut. Chinese and Japanese chestnut trees, which tend to better **survive** blight, don't grow as tall as American chestnut trees, making it hard for them to compete for sunlight in crowded forests." (WSJ)

Against this backdrop, I investigate what roles chestnut trees perform in the corpus and whether agency is attributed to them. In the vast majority of concordance lines *chestnut* is the grammatical object of the sentence and a human agent the subject. Most of the ver-

bal collocates with *chestnut* as object such as *modify*, *engineer*, *revive*, *plant*, *breed*, *harvest*, *save*, *bring back*, *messing with* and *protect* highlight the positioning of the tree as a victim or a passive object of human intervention rather than an agentive participant in ecological processes. The tree is acted upon – engineered, planted, protected – rather than acting itself. This framing reinforces the dominant discourse in which plants, unlike many animals, are more often backgrounded as recipients of human action than foregrounded as active shapers of ecosystems.

Only in a few cases does *chestnut* function as the subject of the clause and in these cases the verbs tend to be less dynamic; the following verbs indicate basic existence (Sealey 2018, 236) or stature: *live*, *die*, *tower*, *sprout*. However, a few verbs also portray the tree's resilience and active part in its 'resurrection': *fight* and *withstand* that appear in a verb phrase and a subclause. Even though these active verbs depict this deciduous tree species as an agent, this capacity is enabled solely through human intervention (who *grow* or *help* the tree to *fight* or *withstand*). Overall, the tree is construed as a victim reliant on humans as saviors, while the original human responsibility for its demise – introducing the devastating fungus – is backgrounded. There are no references to tree characteristics that forestry scientist Suzanne Simard (2018) has recognized in her research which discusses cognitive qualities of trees such as perception, learning, and memory (these do appear in fictional texts, as Poole [2002] shows).

Since the trees are overwhelmingly represented as objects of actions (being saved, being decimated, being genetically engineered) and the concordances show only a few instances of trees initiating processes or having agency, a human-centered framing is reproduced here in which trees are usually cast as static background elements and rather recipients than initiators of actions. Although trees clearly *do* things and play an active role in various ecosystems, the corpus rarely encodes such agency. Instead, the language predominantly reinforces an anthropocentric discourse where trees exist as conditions, resources, or victims, not as agents with intrinsic value, with the American chestnut further framed as an object of human intervention through genetic engineering efforts to resurrect it. Lastly, the fact that an already extinct species receives such great public attention, while other endangered plant species that could still be saved from extinction – or have not even been scientifically described – remain mostly overlooked. This disparity illustrates how extinction is fundamentally a cultural phenomenon (Heise 2016, 5), where cultural values heavily influence conservation priorities and policies – determining which species is considered important enough to be protected or even resurrected.

6 Conclusion

This study aimed to examine how species extinction and endangerment are represented in news discourse with the help of the corpus methods keyword and concordance analysis. The keyword analysis revealed that biodiversity discourse continues to be dominated by references to charismatic megafauna, since most keywords refer to mammalian and bird species. Invertebrates, plants and fungi remain largely marginal, although WP (*Washington Post*) presented a somewhat more balanced range of species than WSJ (*Wall Street Journal*), which foregrounded economic and scientific terminology alongside a nar-

rower taxonomic focus. This tendency reflects what scholars have termed *plant blindness*, a tendency to disregard plants as mere background.

The concordance analyses shed further light on the way plants and fungi are discursively constructed. *Fungus* emerged primarily in negative contexts, salient only when threatening human interests through disease, decay, or economic damage (Cole 2024, 557). Even though *mycorrhiza* appears in the concordance lines, but is only featured in a single article, the discourse predominantly disregards the crucial symbiotic role of fungi within ecosystems. Plants, in contrast, were represented through a narrower set of associations: *milkweed* was almost exclusively portrayed in terms of its instrumental value as food source for monarch butterflies, while *chestnut* appeared overwhelmingly in reference to its attempted revival through genetic engineering. The verbal collocates represent the tree as a victim or as a recipient of human intervention efforts rather than an active agent in its own right. Overall, these findings indicate that media discourse in the two newspapers studied features aspects of beneficial and destructive discourses. On the one hand, the attention to conservation efforts and pollinator-friendly initiatives represents an attempt to raise awareness for ecological issues. On the other hand, such representations often remain anthropocentric and align with what Stibbe (2021, 25–26) refers to as ambivalent discourses: Fungi and plants are primarily portrayed as resources, solutions are framed as small, individualized actions rather than systemic change. When non-human organisms are present in the discourse, but appear only faintly, as objects acted upon rather than as active participants, this results in the narrative of a human-centered world and disregards the intertwined relations of all living beings. Even if a specific tree species is mentioned and the reader can (more or less vividly) imagine it, the word *chestnut* still does not convey "the myriad of shapes of the actual trees, their colours, smells, textures or the complexity of their forms, so there is (as always) some erasure" (Stibbe 2021, 151).

In order to reject these anthropocentric tendencies, it is vital to first detect and understand how inherently speciesist our language system is and on how many levels our language reproduces the lexical distinction human vs. nonhuman (Nübling 2022, 67). To recognize that humans are one of many living beings in a multispecies world, Rob Nixon (2024, 367) emphasizes that trees and other non-humans are "lifely actors, shaping Earth's life systems", evolutionary seniors and potential guides. Bringing such perspectives into mainstream media could help reorient public discourse toward a recognition of shared vulnerability and interconnectivity.

This is why Gabriele Dürbeck et al. (2023, 9) emphasize the need for innovative forms of writing and storytelling such as *The Overstory* that cultivate "a consciousness of interconnectivity between humans, trees, and the natural world that is beneficial for the achievement of more sustainable ways of being" (Poole 2022, 106) in order to change our understanding of the power (*Wirkmächtigkeit*) of animals, plants, mycorrhizal symbioses etc. and foreground multispecies agencies. Media texts can learn from the more diverse resources available in fiction to expand and deepen their representations of nature.

Lastly, I would like to point out several limitations. Since the corpus analyzed here was small and specialized, the findings cannot be generalized to the broader discourse of biodiversity in US media. Keyword analysis as a method is also limited by its comparative design – as mammals, e. g. *elephant* and *wolf* may be highly salient in the reference corpus

(as well as in my target corpus). As a result, my corpus may in fact include a greater overall representation of mammalian species, but this would not be fully captured through keyword analysis. Another limitation arises from the corpus selection, where specifically articles on species loss were filtered out. Due to the specific topical criterion there may be a potential bias toward passive portrayals of plants and fungi as victims of external threats. Consequently, the observed victimization could stem as much from this selection as from ideological framings. Future research could assess the corpus design risk and broaden the corpus to strengthen claims about discursive bias. Additionally, the mostly negative connotations of fungi can be explained by the fact that fungal diseases make up one of the causes of extinction. However, this one-sided framing disregards that key drivers of species loss are, after all, anthropogenic, such as habitat destruction and climate change (Lughadha et al. 2020, 396), as well as ignoring the fact that fungi can also positively impact biodiversity.

Despite these limitations, the study demonstrates how corpus-assisted discourse analysis can help uncover systemic absences and marginalizations in biodiversity discourse. By drawing attention to these patterns, ecolinguistics can contribute to "reminding" (Stibbe 2021) – bringing marginalized species and erased interdependencies back into consciousness. Media representations play a crucial role in communicating scientific findings to the public and shaping how people think about the environment. A more balanced portrayal of the vast richness of non-human life – not merely as resources but as co-inhabitants with agency – is vital in fostering the cultural and linguistic conditions for ecological sustainability and multispecies justice.

Acknowledgement: This work was supported by the Estonian Research Council under Grant PRG1504 and Grant PRG3262.

7 Bibliography

Baker, Paul. *Using Corpora in Discourse Analysis*. Continuum, 2006.

Baker, Paul, Costas Gabrielatos, and Majid Khosravinik. "A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press." *Discourse & Society*, vol. 19, no. 3, 2008, pp. 273–306. https://doi.org/10.1177/0957926508088962.

Blades, Byrony. "What's in a Name? An Evidence-based Approach to Understanding the Implications of Vernacular Name on Conservation of the Painted Dog (Lycaon Pictus)." *Language & Ecology*, 2020, pp. 1–27.

Bray, Carly. "What Do We Mean by 'Representation'? Towards a Systematic Corpus-Assisted Critical Discourse Analysis of First Nations People(s) in Australian Print News." *Critical Approaches to Discourse Analysis across Disciplines*, vol. 16, no. 2, 2024, pp. 18–50.

Buzbee, Sally. "Introducing the Post's Expanded Climate Coverage." *The Washington Post*, 28 Nov. 2022, https://www.washingtonpost.com/climate-environment/2022/11/28/introducing-posts-expanded-climate-coverage/. Accessed 26 Sept. 2025.

Campbell, Sueellen. "Elegy for the Washinton Post Climate Team." *Yale Climate Connections*, 17 Feb. 2026, https://yaleclimateconnections.org/2026/02/elegy-for-the-washington-post-climate-team/. Accessed 1 March 2026.

Cole, Hannah Rachel. "'Spurned Growth'. Reading for Rot in The Kingdom of This World." Environmental Humanities, vol. 16, no. 3, 2024, pp. 554–570.

Dürbeck, Gabriele, Urte Stobbe, and Evi Zemanek. "Netzwerke des Lebendigen: Multispecies Agencies und Formexperimente in hybriden Genres. Einleitung." *Transpositiones*, vol 2, no. 2, 2023, pp. 9–12.

Fairclough, Norman, and Ruth Wodak. *Critical Discourse Analysis. Discourse Studies: A Multidisciplinary Introduction*, ed. by Teun Van Dijk, Sage, 1997, pp. 258–284.

Fill, Alwin, and Peter Mühlhäusler, eds. *The Ecolinguistics Reader. Language, Ecology and Environment*. Continuum, 2001.

Finzi, Joshua. "Fungal Rhizomes: Mycorrhizae and Posthumanism." *Green Letters: Studies in Ecocriticism*, vol. 28, no. 1–2, 2024, pp. 1–16.

Gillings, Mathew, Gerlinde Mautner, and Paul Baker. *Corpus-Assisted Discourse Studies*. Cambridge University Press, 2023.

Goatly, Andrew. *Critical Reading and Writing: An Introductory Coursebook*. Routledge, 2002.

Grandoni, Dino. "Gene Editing Could Revive a Nearly Lost Tree. Not Everyone Is on Board." *The Washington Post*, 30 August 2022, https://www.washingtonpost.com/climate-solutions/2022/08/30/american-chestnut-blight-gene-editing/. Accessed 28 Sept. 2025.

Halliday, Michael. "New Ways of Meaning: The Challenge to Applied Linguistics." 1990. *The Ecolinguistics Reader: Language, Ecology and Environment*, ed. by Alwin Fill and Peter Mühlhäusler. Continuum, 2001, pp. 175–202. (Reprinted from "New Ways of Meaning: The Challenge to applied linguistics," *Journal of Applied Linguistics*, vol. 6, 1990, pp. 7–36.)

Haugen, Einar. "The Ecology of Language." 1972. *The Ecolinguistics Reader: Language, Ecology and Environment*, ed. by Alwin Fill, and Peter Mühlhäusler, Continuum, 2001, pp. 57–66. (Reprinted from *The Ecology of Language: Essays by Einar Haugen*, Stanford University Press, 1972, pp. 325–329.)

Heise, Ursula. *Nach der Natur. Das Artensterben und die moderne Kultur*. Suhrkamp, 2010.

Heise, Ursula. *Imagining Extinction. The Cultural Meanings of Endangered Species*. The University of Chicago Press, 2016.

Lakoff, George. *Don't Think of an Elephant! Know your Values and Frame the Debate: The Essential Guide for Progressives*. Chelsea Green, 2014.

Lughadha, Eimar Nic, Steven P. Bachman, and Tarciso C. C. Leão. "Extinction Risk and Threats to Plants and Fungi." *Plants, People, Planet*, vol. 2, no. 5, 2020, pp. 389–408.

Nixon, Rob. "The Less Selfish Gene. Forest Altruism, Neoliberalism, and the Tree of Life." *Environmental Humanities*, vol. 12, no. 2, 2021, pp. 348–371.

Nübling, Damaris. "Linguistische Zugänge zur Tier/Mensch-Grenze." *Mensch – Tier – Maschine: Sprachliche Praktiken an und jenseits der Außengrenze des Humanen*, ed. by Miriam Lind, transcript, 2022, pp. 27–76.

Ponton, Douglas Mark. *Exploring Ecolinguistics. Ecological Principles and Narrative Practices*. Bloomsbury, 2024.

Poole, Robert. *Corpus-Assisted Ecolinguistics*. Bloomsbury, 2022.

Powers, Richard. *The Overstory*. Penguin, 2018.

Rayson, Paul, and Amanda Potts. "Analysing Keyword Lists." *A Practical Handbook of Corpus Linguistics*, ed. by Magali Paquot and Stefan Th. Gries, Springer International, 2021, pp. 119–139.

Scott, Mike, and Christopher Tribble. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Benjamins, 2006.

Sealey, Alison. "Animals, Animacy and Anthropocentrism." *International Journal of Language and Culture*, vol. 5, no. 2, 2018, pp. 224–247.

"Seven of Nine Planetary Boundaries Now Breached. Ocean Acidification Joins the Danger Zone." *Potsdam Institute for Climate Impact Research*, 24. Sept. 2025. https://www.pik-potsdam.de/en/news/latest-news/seven-of-nine-planetary-boundaries-now-breached-2013-ocean-acidification-joins-the-danger-zone. Accessed 26 Sept. 2025.

Sheldrake, Merlin. *Entangled Life: How Fungi Make Our Worlds, Change Our Minds, and Shape Our Futures*. Penguin, 2020.

Simard, Suzanne W. "Mycorrhizal Networks Facilitate Tree Communication, Learning, and Memory." *Memory and Learning in Plants*, ed. by Frantisek Baluska, Monica Gagliano, and Guenther Witzany. Springer International, 2018, pp. 191–213.

Stibbe, Arran. *Animals Erased. Discourse, Ecology, and Reconnection with the Natural World*. Wesleyan University Press, 2012.

Stibbe, Arran. "Ecolinguistics and Erasure: Restoring the Natural World to Consciousness." *Contemporary Critical Discourse Studies*, ed. by Christopher Hart and Piotr Cap, Bloomsbury Academic, 2014, pp. 583–602.

Stibbe, Arran. "Corpus Discourse Analysis and Ecology." *The Routledge Handbook of Critical Discourse Studies*, ed. by John Flowerdew, and John E. Richardson, Routledge, 2017, pp. 497–509.

Stibbe, Arran. *Ecolinguistics. Language, Ecology and the Stories We Live By*, 2nd ed., Routledge, 2021.

Subtirelu, Nicholas Close, and Paul Baker. "Corpus-based Approaches." *The Routledge Handbook of Critical Discourse Studies*, ed. by John Flowerdew, and John E. Richardson. Routledge, 2018, pp. 106–119.

Tsing, Anna. "Unruly Edges: Mushrooms as Companion Species." *Environmental Humanities*, vol. 1, 2012, pp. 141–154.

United Nations (UN). *The 17 Goals*. https://sdgs.un.org/goals. Accessed 20 Feb. 2026.

United Nations (UN). *Transforming our World: The 2030 Agenda for Sustainable Development*, 2015, https://www.un.org/pga/wp-content/uploads/sites/3/2015/08/120815_outcome-document-of-Summit-for-adoption-of-the-post-2015-development-agenda.pdf. Accessed 20 Feb. 2026.

Widdowson, Henry George. *Text, Context, Pretext. Critical Issues in Critical Discourse Analysis*. Blackwell, 2014.

WWF. *Monarch Butterfly*. https://www.worldwildlife.org/species/monarch-butterfly/. Accessed 25 Sept. 2025.

Wulff, Stefanie, and Paul Baker. "Analyzing Concordances." *A Practical Handbook of Corpus Linguistics*, ed. by Magali Paquot, and Stefan Th. Gries, Springer, 2020, pp. 161–180.

Appendix

Table 3: Top 50 keywords in The Washington Post corpus

	Frequency (WP)	Frequency (reference corpus)	Relative frequency[a] (WP)	Relative frequency (reference corpus)	ARF[b] (WP)	Score[c]
vaquita	54	7	200,6115	0,00694	4,09393	200,22
bison	105	981	390,07791	0,97203	8,83899	198,31
firefly	71	682	263,76697	0,67576	4,6304	158
maui	52	343	193,18144	0,33986	6,77398	144,93
amphibian	61	833	226,61668	0,82538	16,30073	124,7
biologist	126	3098	468,09348	3,06966	63,48239	115,27
ecosystem	178	4842	661,2749	4,79771	91,0842	114,23
habitat	295	9615	1095,93311	9,52705	139,38255	104,2
smithsonian	56	1065	208,04155	1,05526	18,48449	101,71
xerces	26	4	96,59072	0,00396	6,76124	97,205
pollinator	40	670	148,6011	0,66387	13,09432	89,911
pup	89	2733	330,63745	2,708	15,98187	89,438
coyote	38	602	141,17105	0,59649	7,5085	89,052
milkweed	25	70	92,87569	0,06936	6,76501	87,787
mammal	111	3808	412,36807	3,77317	48,56729	86,602
guam	36	596	133,741	0,59055	5,10617	84,714
woodpecker	40	792	148,6011	0,78476	4,6112	83,822
condor	34	528	126,31094	0,52317	4,16281	83,583
whale	266	11448	988,19733	11,34329	33,19133	80,141
invasive	79	2727	293,48718	2,70206	31,17267	79,547
fungus	91	3302	338,0675	3,2718	19,14337	79,374
nonnative	22	47	81,73061	0,04657	8,9253	79,049
ecologist	46	1202	170,89127	1,19101	22,72674	78,453
ferret	45	1254	167,17624	1,24253	4,83571	74,994
salwan	20	25	74,30055	0,02477	5,28838	73,48
die-off	21	77	78,01558	0,0763	9,33836	73,414
imperil	32	737	118,88088	0,73026	17,20623	69,285
black-footed	19	37	70,58553	0,03666	4,34333	69,054
rhino	125	5822	464,37845	5,76875	10,59448	68,754
conservation	373	19744	1385,70532	19,5634	185,97865	67,436
turtle	109	5109	404,93802	5,06227	18,65152	66,961
conservationist	76	3312	282,3421	3,28171	35,71793	66,175
chestnut	79	3508	293,48718	3,47591	4,02712	65,794
hatchling	21	220	78,01558	0,21799	6,42769	64,874
iucn	23	355	85,44563	0,35175	8,80686	63,951
zoo	279	15500	1036,49268	15,35822	56,87561	63,423
salamander	24	430	89,16067	0,42607	4,64945	63,223
wildlife	446	26672	1656,90234	26,42804	201,5161	60,446
organism	53	2326	196,89647	2,30472	23,29695	59,883

	Frequency (WP)	Frequency (reference corpus)	Relative frequency[a] (WP)	Relative frequency (reference corpus)	ARF[b] (WP)	Score[c]
alligator	35	1301	130,02597	1,2891	7,79628	57,239
grizzly	36	1441	133,741	1,42782	6,33594	55,499
subspecies	17	204	63,15547	0,20213	7,35836	53,368
veterinarian	29	1058	107,7358	1,04832	10,68381	53,085
geologic	15	70	55,72541	0,06936	4,06636	53,046
cheetah	32	1338	118,88088	1,32576	4,23645	51,545
grassland	34	1538	126,31094	1,52393	12,06911	50,442
insect	133	8992	494,09866	8,90975	46,96872	49,961
frog	88	5633	326,92242	5,58148	16,2398	49,825
specimen	63	3842	234,04674	3,80686	20,12843	48,898
paleontologist	15	168	55,72541	0,16646	4,35875	48,63
pollinate	20	567	74,30055	0,56181	10,82796	48,214

a) Frequency per million is the number of occurrences of a word per million tokens
b) Average Reduced Frequency (ARF) is a modified frequency that 'discounts' multiple occurrences of a word that appear in the same document
c) The keyness score calculated using the simple math method

Table 4: Top 50 keywords in The Wall Street Journal corpus

	Frequency (WSJ)	Frequency (reference corpus)	Relative frequency[a] (WSJ)	Relative frequency (reference corpus)	ARF[b] (WSJ)	Score[c]
manatee	58	288	321,37057	0,28537	3,83717	250,8
lobsterman	39	51	216,09402	0,05053	5,38006	206,65
koala	57	637	315,82971	0,63117	4,45598	194,23
bison	49	981	271,50275	0,97203	5,11256	138,18
whale	286	11448	1584,68945	11,34329	54,5341	128,47
cop15	22	23	121,89919	0,02279	5,22911	120,16
woodpecker	37	792	205,01227	0,78476	4,52996	115,43
greenhouse-gas	22	131	121,89919	0,1298	7,32797	108,78
grizzly	45	1441	249,33925	1,42782	4,62709	103,11
fossilize	19	72	105,27657	0,07134	5,3906	99,2
nature-related	16	9	88,65395	0,00892	4,77865	88,862
condor	21	528	116,35831	0,52317	3,13567	77,049
birder	20	460	110,81744	0,45579	4,50067	76,809
ecosystem	76	4842	421,10629	4,79771	32,00389	72,806
padlock	23	784	127,44006	0,77683	3,15369	72,286
audubon	18	493	99,7357	0,48849	7,27567	67,676
pollination	16	401	88,65395	0,39733	5,31579	64,161
tasmanian	23	1022	127,44006	1,01265	3,22684	63,816

	Frequency (WSJ)	Frequency (reference corpus)	Relative frequency[a] (WSJ)	Relative frequency (reference corpus)	ARF[b] (WSJ)	Score[c]
deforestation	28	1477	155,14442	1,46349	12,34017	63,383
coyote	18	602	99,7357	0,59649	5,08687	63,098
traps	45	3173	249,33925	3,14398	7,19394	60,41
maine	46	3385	254,88013	3,35404	10,25289	58,768
fossil-fuel	13	268	72,03134	0,26555	5,57236	57,707
habitat	105	9615	581,79156	9,52705	43,92994	55,361
bald	48	3886	265,96185	3,85046	7,37291	55,039
pollinator	16	670	88,65395	0,66387	3,22518	53,883
wind-farm	10	54	55,40872	0,05351	3,18695	53,544
turbine	70	6359	387,86105	6,30084	14,61196	53,263
wildfire	40	3322	221,63489	3,29161	8,69394	51,877
rainforest	45	3907	249,33925	3,87126	16,90376	51,391
conservancy	18	1057	99,7357	1,04733	6,59208	49,203
birding	11	290	60,9496	0,28735	3,08947	48,122
lobster	53	5193	293,66623	5,1455	12,32496	47,948
biologist	35	3098	193,93053	3,06966	15,26735	47,898
fishery	51	4991	282,58447	4,94535	16,17745	47,699
oceanic	16	899	88,65395	0,89078	8,65111	47,416
conservation	175	19744	969,65265	19,5634	75,77764	47,203
ecologist	18	1202	99,7357	1,19101	10,34693	45,977
carbon	230	27286	1274,40063	27,03642	47,73195	45,491
eagle	92	10406	509,76025	10,31082	12,69765	45,157
humpback	13	625	72,03134	0,61928	7,70304	45,101
climate-change	11	418	60,9496	0,41418	6,32395	43,806
mammal	37	3808	205,01227	3,77317	19,30034	43,16
panda	41	4383	227,17577	4,34291	3,50322	42,706
insect	73	8992	404,48367	8,90975	14,68918	40,918
migratory	17	1343	94,19483	1,33072	4,1114	40,844
nestlé	12	665	66,49046	0,65892	3,10939	40,683
wildlife	200	26672	1108,17444	26,42804	78,52978	40,439
nonprofit	38	4283	210,55315	4,24382	22,03067	40,343
marsupial	9	281	49,86785	0,27843	5,036	39,789

a) Frequency per million is the number of occurrences of a word per million tokens

b) Average Reduced Frequency (ARF) is a modified frequency that 'discounts' multiple occurrences of a word that appear in the same document

c) The keyness score calculated using the simple math method