
Sentiment and Structure in Reintroduction Science: A Meta-Analysis Using NLP Tools

Alfonso AWADALLA CARREÑO
École Polytechnique / ENSAE
MSc&T Data & Economics for Public Policy
May 4, 2025

Abstract

This study presents an automated text-mining pipeline applied to over 4,000 scientific abstracts from the species reintroduction literature (1987–2016). Using lexicon-based sentiment models and custom scripts in R, I extract and standardise taxonomic names, IUCN Red List categories, and geographic references. Sentiment is quantified using an ensemble of five models, revealing a slight positive trend over time—though this should be interpreted with caution due to sample variability. The results also highlight persistent biases in conservation science: mammals and birds dominate the literature, high-income countries are overrepresented, and less-threatened species tend to receive more coverage. By combining sentiment analysis with structured information extraction, this approach enables scalable meta-analysis of conservation texts and offers new insights into how scientific attention is distributed across taxa, threat levels, and regions.

1 Introduction

1.1 Motivation & Context

Although they may not be of interest to you, dear reader, the motivations behind this study are primarily personal. As part of a course designed to apply newly acquired knowledge in machine learning for natural language processing (NLP), I was free to choose a topic I liked, and being deeply passionate about ecology and conservation, I naturally turned to this one, which also gave me the opportunity to learn a lot. Since this report is not intended for a niche audience, let me introduce its general and scientific context.

In this study, we will talk about species, taxonomy, IUCN threat categories, and finally the success or failure of a species reintroduction. The notion of species is a blurry concept — Charles Darwin illustrates this well in the Chapter II of his major work *On the Origin of Species* [1859], in which he explains — in a context where genetics did not yet exist — that two organisms are considered distinct species if they are 'different enough'. These differences were generally based on the organisms' physical appearance and were, ultimately, subjective. Before the species level, there is also the subspecies or even the variety, and so defining what counts as 'enough' difference to classify two close organisms separately was a complicated task.

Darwin's argument — which was later validated by advances in genetics — was that every closely related species once had a common ancestor and, at some point in history, it was first a variety, then a subspecies of another, before becoming different enough to be considered a species. And all of this process happens through a long process of natural selection (environmental conditions, interactions with other organisms) acting on individual variation — which, as we now know, arises from random genetic mutations.

Taxonomy is the system that we use to classify living things and provides a structured way to describe biodiversity. It was first introduced by Carl Linnaeus in the 18th century, long before anyone knew what DNA was, and it is still more or less the structure we use today. The idea is to organise life from the broadest groups down to the most specific, which gives us the now-familiar ladder of Kingdom, Phylum, Class, Order, Family, Genus, and Species. Since then, the system has evolved, especially with the arrival of evolutionary theory and genetics, which made it possible to classify organisms not only by how they look, but also by how they're related through common ancestry [Mayr, 1982, Woese et al., 1990]. Figure 1 helps visualise this and shows how this works for humans. In this study, we will stick to this classic hierarchy and focus mainly on the order or species level.

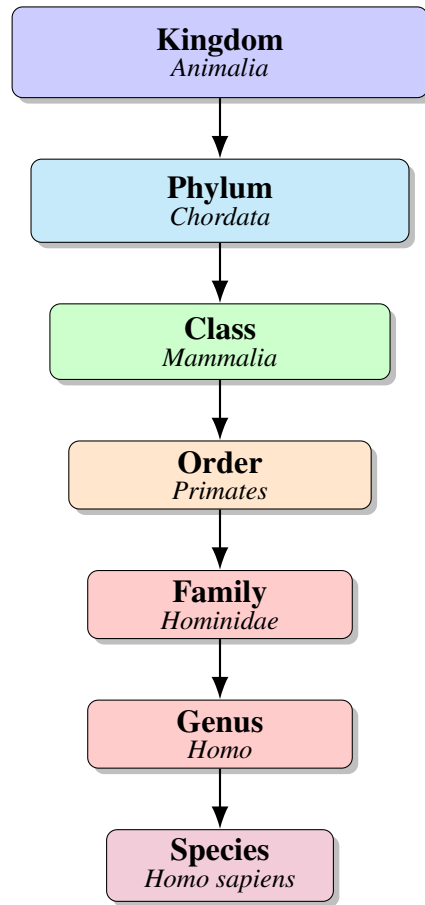


Figure 1: Taxonomic hierarchy of the human — from general to specific

Now that we have clarified how species are defined and classified, it is important to introduce how their conservation status is evaluated. The most widely used framework is the IUCN Red List of Threatened Species, which categorises species based on a combination of ecological and demographic criteria — including population trends, geographic distribution, and habitat fragmentation [IUCN, 2012]. These categories range from “Least Concern” to “Critically Endangered” and “Extinct” and when a species is considered at risk or locally extinct, one strategy sometimes employed is reintroduction: the release of individuals into parts of their historical habitat. These initiatives, often based on captive breeding or translocations from more stable populations, are intended to reestablish viable, self-sustaining populations in the wild. But reintroduction can be complex and success depends on various factors, from habitat suitability, to interactions with other organisms or public support, making failure not so uncommon [Seddon et al., 2014]. In this study, we explore the trends of these reintroduction attempts over time.

1.2 Problem Statement

As a student, I often feel overwhelmed by the huge amount of research papers out there — papers that I know are interesting, but far too long and numerous. Of course, some are more rigorous than others and manage to stay up to date, but I do often feel like I am missing out on a lot of valuable 'knowledge'. Meta-analysis is a method that consists in systematically combining and analysing the results of multiple studies to identify patterns, trends, and overall findings.

Bernard Lahire [2023], in his recent book, argues that meta-analysis is not merely a way of synthesising research, but a tool that allows us to take a step back from science itself — to understand its structures and the biases in what is studied, and how. Often, research on a topic becomes too specific and is conducted in isolation, losing sight of the main focus of the science and what it should be studying. Although Lahire's book focuses on sociology, his ideas and critiques, in my view, are applicable to any field, including conservation biology.

Trisha Greenhalgh [1997], in a great article titled "How to read a paper: Papers that summarise other papers", contrasts traditional narrative reviews — like many academic literature reviews (and very probably the one I am writing here) — which tend to be overly selective. Instead, she advocates for systematic reviews that rely on replicable and transparent methods, scalable in scope, and which reduce selection bias while, as Lahire suggests, also offering insights into the structure of research itself.

With this in mind, recent methods in Natural Language Processing (NLP) and sentiment analysis could allow for a form of meta-analysis of research in conservation biology and species reintroduction, using a rigorous and scalable methodology as Greenhalgh suggests. Addressing these challenges and shedding light on the methodological biases involved is the main goal of this study.

2 Literature Review & Contributions

I want to begin this section by highlighting the three papers that served as my main references. This study can, in a way, be seen as a mix of both.

Van Houtan et al. [2020], first of all, provided the dataset I used in this study: over 4,300 scientific abstracts from conservation biology research, which I describe in more detail in the Data section. Second, the authors performed sentiment analysis on these abstracts — a machine learning algorithm that quantifies the 'sentiment' score (positive or negative) of the texts. I drew inspiration from their methodology to carry out a similar analysis on the abstracts, with a few slight modifications, as outlined in the Experimental Setup & Methodology section.

The second reference is the paper by Evans et al. [2022], which focuses on extracting specific variables from abstracts in the reintroduction biology literature. They successfully extract taxonomic information (species), the IUCN threat status of the species studied, and the geographical context of the research. The last key reference is the study by Lennox et al. [2020], which combines both sentiment analysis — using a slightly different methodology than Van Houtan et al. [2020] — and systematic information extraction (taxa and Red List categories).

Although the scripts and methods used in this study are my own work, it's important to acknowledge where these ideas came from — and also to note that many of the figures I generated were inspired by the clear and well-designed visuals in these two papers.

2.1 Sentiment Analysis in Scientific Texts

Language, and the way people express themselves, is a fascinating topic, but it also raises many questions, especially when we consider that each person interprets and communicates things differently depending on the context, whether cultural or domain-specific. Sentiment analysis starts by creating a lexicon or dictionary of words, each assigned a "sentiment score" — for instance, the word 'success' might receive a positive score, while 'failure' would be assigned a negative one. By extracting words from a text and summing their associated scores, it is then possible to estimate the overall sentiment of the text.

Bender & Friedman [2018] highlights several limitations of this method. First, they argue that the interpretation of language can vary across individuals and over time. A common issue in studies

using NLP is that sentiment lexicons are often trained and developed in very specific contexts (such as movie or restaurant reviews), meaning that a word might carry a certain value in one context but a different one in another. This is why it is crucial, whenever possible, to develop a lexicon that is specific to the subject of study. Unfortunately, that is not always feasible — as is the case here — and I will return to this point in the discussion of limitations.

Van Houtan et al. [2020] applied sentiment analysis to over 4,000 abstracts to measure how positively or negatively conservation science describes reintroduction outcomes. To address the bias arising from using lexicons coming from other contexts, they used 5 different models and computed an average of them to get a more robust sentiment score. In order to analyse sentiment score across species, they simply scraped the name of famous species like "shark", "otter" or "lynx" and then selected the abstracts containing those strings. This poses a problem if the objective is to do a meta analysis of the science and compare all the species mentioned and studied.

Lennox et al. [2020] conducted a similar study on a larger set of conservation abstracts. They found that sentiment varied across taxonomic groups and threat levels — with more positive tone often associated with charismatic or less threatened species. Methodologically, they used an ensemble of four sentiment models, three of which overlapped with those used by Van Houtan et al. [2020], and a different one. Unlike Van Houtan et al. [2020], they standardised the scores from each model before averaging, aiming for a more balanced ensemble sentiment score.

2.2 Information Extraction in Ecology

While sentiment tells us how things are written, it does not tell us what is being written about — and that is where information extraction becomes essential. Extracting structured information from raw text can reveal patterns that are otherwise buried in language: for example, which species are being studied, whether they are threatened or not, and where the study took place. But identifying this information reliably in thousands of abstracts is not as easy as it sounds.

One widely used approach for extracting taxonomic names is the SPECIES tagger introduced by Pafilis et al. [2013]. It is a dictionary-based software designed to identify scientific names in large volumes of text, especially in biomedical and biodiversity contexts. Its main advantage lies in speed and accuracy, making it particularly suitable for large-scale applications. However, like other dictionary-based tools, it can struggle with ambiguous names or taxonomic synonyms, which are common in ecological literature.

Evans et al. [2022] developed a structured pipeline to extract species names, geographical information, and IUCN threat categories from over 3,700 abstracts in the reintroduction literature. For taxonomic data, they used the Global Names Finder (GNF), part of the GNRD suite, to detect scientific names in abstracts and titles, and then retrieved higher-level classifications using the `taxize` package in R. To extract geographic entities, they relied on the CLIFF-CLAVIN geoparser, which identifies and geolocates place names, and then categorised them by country. Finally, to match species mentions to their Red List status, they used the `rredlist` package. This pipeline allowed them to generate a clean dataset that revealed strong biases in the literature — toward mammals and birds, and toward species that were not actually considered threatened — illustrating how well-designed information extraction methods can offer new insights into where scientific attention is directed.

Lennox et al. [2020], although primarily concerned with sentiment analysis, also extracted taxonomic and threat-level information to compare emotional tone across species. Their method was simpler — relying on species mentions already present in the text, which they linked to taxonomic data using the `taxize` package, and to IUCN Red List categories using the `rredlist` package.

More generally, Thessen et al. [2012] reviewed the broader use of natural language processing in biodiversity science and highlighted some of the key challenges of information extraction in this domain, such as ambiguous scientific names, nested entities, and the lack of annotated corpus for training or evaluation. Their work reinforces the idea that domain-specific adaptation is crucial when applying NLP to ecological texts, and that even simple extraction tasks often require carefully designed workflows.

2.3 Contributions of this Study

This study uses the abstract dataset compiled by Van Houtan et al. [2020], who did not conduct any systematic information extraction from the texts. My contribution is to apply such extraction methods to this dataset, thus enriching the analysis and potentially uncovering new insights.

Secondly, I deliberately chose not to use external tools/software like those mentioned in the previous subsection — not only to avoid unnecessary complexity, but also to keep the methodology as simple and transparent as possible. While this decision comes with certain limitations, I developed innovative techniques (all scripted in RStudio), particularly for retrieving geographical information from the text.

All the code is clearly structured, well-commented, and publicly available on my GitHub profile for reproducibility and further use.¹

3 Data

3.1 Data Description

The dataset used in this study originates from the work of Van Houtan et al. [2020], who compiled and analysed 4,313 abstracts on species reintroductions and translocations published between 1987 and 2016. These abstracts were retrieved from the Web of Science using Boolean queries that targeted key terms such as *species*, *conservation*, *population*, and either *reintroduction* or *translocation*, while filtering out studies from molecular biology and other unrelated fields. The final corpus includes over one million words and represents a substantial portion of the published literature on conservation-oriented reintroduction efforts.

To visualise the spatial distribution of reintroduction research and species studied, I used an open-access shapefile of country boundaries from the World Administrative Boundaries dataset, available on OpenDataSoft². This allowed me to generate maps linking abstracts to countries based on geographic mentions extracted in the information extraction pipeline.

3.2 Structure and Content of the Dataset

The CSV file used³ contains one row per article, with variables such as abstract text, publication year, journal name, and metadata extracted by the authors.

As part of my own processing, I extracted additional information on species, geographic mentions, and IUCN categories using the methods described in the next section. Figure 2 illustrates how many abstracts in the original dataset contain usable information for each type of analysis, and how the sample size decreases as more metadata layers are required.

3.3 Limitations

The abstract dataset also comes with several limitations. First, it only includes abstracts, which means the analysis is based on a condensed summary of the research, often omitting key details found in full texts. Second, the dataset is restricted to articles published in English and indexed in the Web of Science, which introduces a linguistic and geographic bias in the corpus. This likely leads to an overrepresentation of studies from high-income, English-speaking countries, and may underrepresent research conducted in the Global South. Finally, the dataset ends in 2016 and therefore does not capture more recent trends in reintroduction literature over the past few years.

¹<https://github.com/alfonsoawd/Project-NLPSentimentAnalysis>

²<https://public.opendatasoft.com/explore/dataset/world-administrative-boundaries/>

³Available here: https://github.com/MBayOtolith/nlp_conservation_lit_review_2018

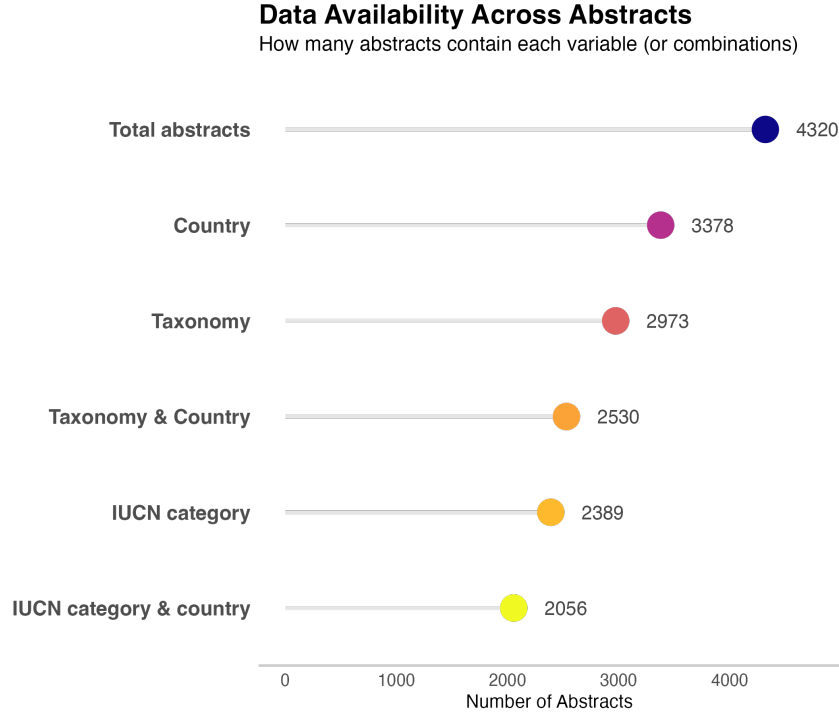


Figure 2: Sample attrition by metadata availability

4 Experimental Setup & Methodology

This section outlines the analytical pipeline developed for this study. While the dataset was originally compiled by Van Houtan et al. [2020], all further data processing, extraction of structured metadata, and sentiment scoring were performed independently. The methodology consists of two main stages: extracting relevant information from abstracts and computing a composite sentiment score for each text.

4.1 Information Extraction Pipeline

To extract structured information from the abstracts, I implemented a pipeline inspired by the work of Evans et al. [2022] and Lennox et al. [2020]. Taxonomic names were first identified using the `scrapenames` function from the `taxize` package in R, which detects scientific names in unstructured text. A key challenge here was that multiple species names could appear within a single abstract. To address this, I selected the first species mentioned, assuming it to be the primary focus of the study — a pattern confirmed through manual checks of several abstracts. Once identified, these names were resolved into higher-level taxonomic ranks (genus, order, class, and kingdom) by querying the NCBI taxonomy database via a free NCBI API⁴, using the `tax_name` function from the `taxize` package.

To determine the conservation status of each species, I used the `rredlist` package to match the extracted names to the latest IUCN Red List categories, which also required the use of a free API from the IUCN website⁵. When species appeared under outdated or synonymous names, synonym resolution functions were applied to ensure the best possible match.

Additionally, I attempted to infer the country of focus for each abstract using a language model API⁶, which responded to a custom prompt I wrote to scan the text and return the most likely geographic reference. These geographic mentions were then normalised and cleaned to align with the shapefile used for mapping and to support further geographic analyses.

⁴NCBI: <https://support.nlm.nih.gov/kbArticle/?pn=KA-05317>

⁵IUCN: <https://api.iucnredlist.org/>

⁶Zhipu AI: <https://bigmodel.cn/>

4.2 Sentiment Scoring Procedure

The second part of the pipeline consisted of computing sentiment scores for each abstract. First, the text was cleaned and preprocessed: all non-English segments were removed, and the abstracts were tokenised and lemmatised to improve matching against lexicons. I then applied five sentiment models: four lexicon-based approaches (Bing, AFINN, NRC, and Syuzhet) [Van Houtan et al., 2020] and one syntactic-contextual model (`sentimentr`) [Lennox et al., 2020]. Each method returned a sentiment score, which was then standardised to z-scores using the classical formula:

$$z_i = \frac{x_i - \mu}{\sigma}$$

where x_i is the original sentiment score for an abstract, μ is the mean sentiment score across all abstracts for that model, and σ is the standard deviation.

I then computed a composite sentiment score by averaging the standardised values from all five models. This ensemble score is meant to balance the strengths and limitations of each individual method and provide a more robust estimate of the abstract’s overall emotional tone.

5 Results and Analysis

5.1 Temporal Sentiment Trends of Abstracts

Figure 3 shows the evolution of average standardised sentiment scores across the reintroduction literature from 1987 to 2016. Because the scores have been standardised, the sign of the score is less informative than its trajectory over time, so interpretation should focus on trends rather than absolute value. Early years shows higher variability, reflecting the low number of abstracts for this period. This uneven distribution of abstracts was not mentioned by Van Houtan et al. [2020], who emphasised only the positive and less variable sentiment over years.

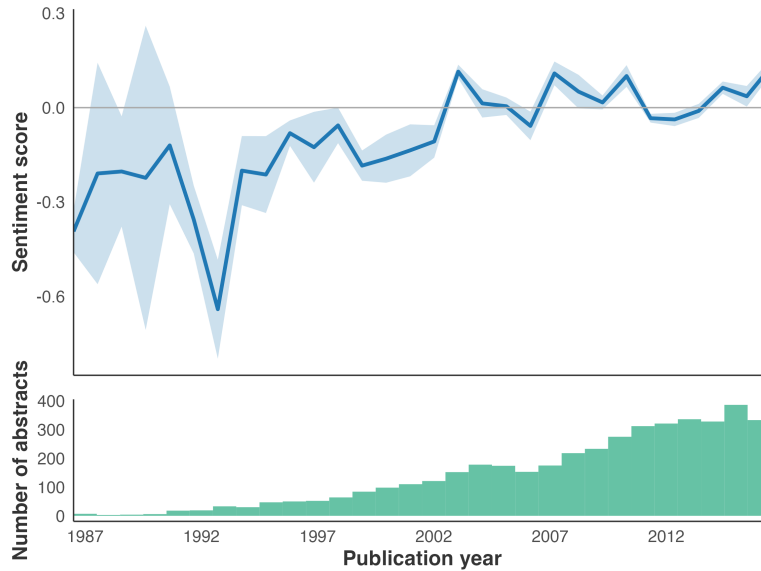


Figure 3: Average standardized sentiment score of abstracts over time (1987–2016).

5.2 Word-Level Sentiment and Lexicon Influence

To better understand what drives the sentiment scores, Figure 4 shows the most influential words in the sentiment lexicons used. For clarity, the AFINN lexicon was chosen (out of the five evaluated)

to generate this figure. Terms like *success*, *increase*, and *threaten* and *loss* dominate the polarity. This aligns with what has been observed by Van Houtan et al. [2020], where the tone of conservation writing often revolves around contrasts between achievement and loss.

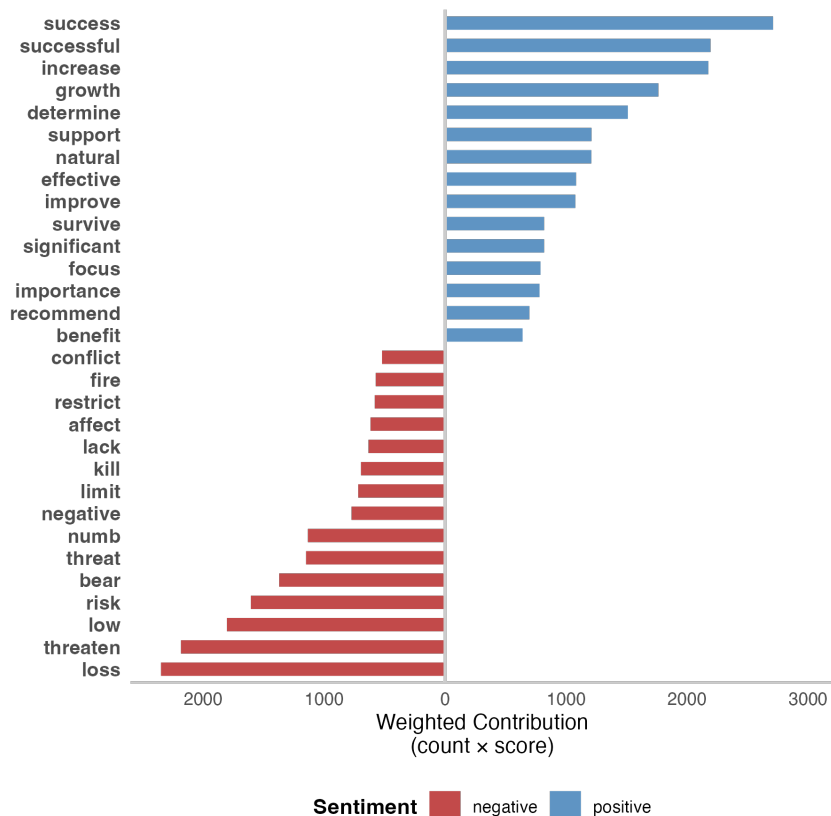


Figure 4: Mirror bar of the 15 words with largest contribution to sentiment score.

5.3 Sentiment Trends by Threat Category

Figure 5 shows the mean standardised sentiment score for each IUCN Red List category. Consistent with Lennox et al. [2020], scores tend to decrease for more threatened categories. However, those categories (e.g., Critically Endangered, Extinct in the Wild) tend to have far fewer abstracts, leading to higher standard deviations and greater uncertainty in their trends.

5.4 Taxonomic Focus and Threat-Level Bias

Figure 6 highlights the top 20 animal orders by frequency and their distribution across IUCN threat levels. Carnivores and hoofed mammals dominate, followed by birds and amphibians. These results coincide with findings from Evans et al. [2022], who point to a strong bias toward mammals and birds, especially those that are well-known or charismatic.

An open question is whether this imbalance in representation reflects the natural proportion of species in each order or a bias toward more publicly interesting taxa. In any case, it is striking to see that a relatively low proportion of the species studied are classified as threatened, again reinforcing the research–relevance gap pointed out by Lennox et al. [2020].

5.5 Sentiment Trends by Species and IUCN Status

Sentiment also varies across species and Red List categories. Figure 7 shows scores for the top three most-mentioned species in each red list category. Some of these trends align closely with the

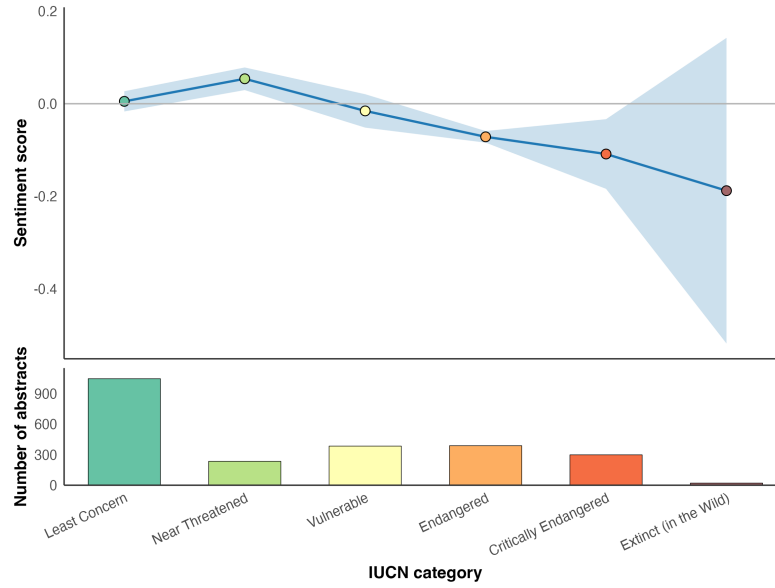


Figure 5: Mean standardised sentiment score by IUCN Red List category.

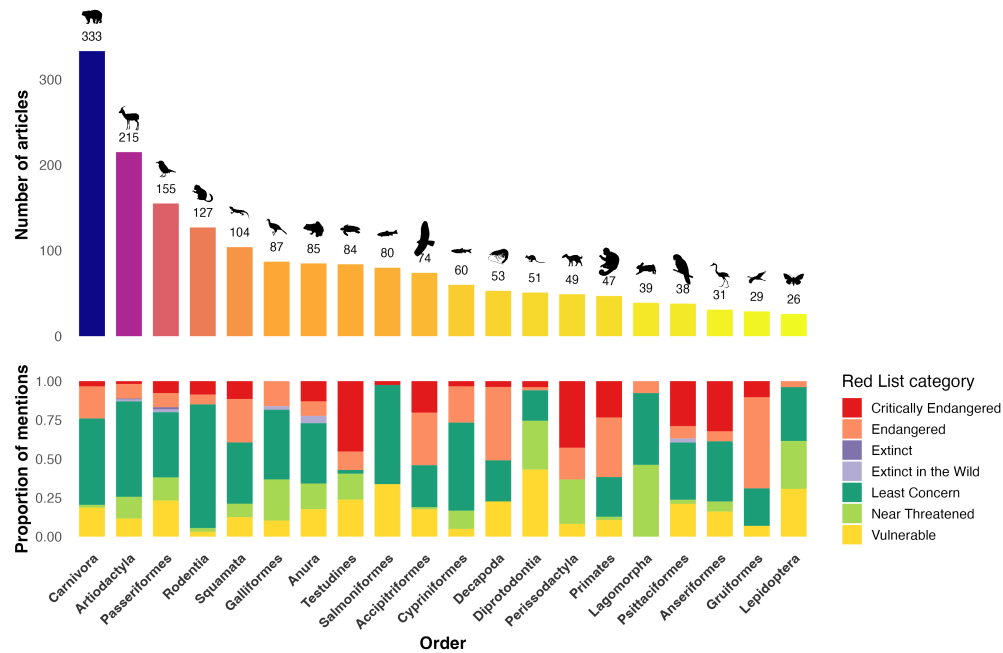


Figure 6: Top 20 most represented animal orders and their IUCN category distribution.
 Note: Silhouettes downloaded from <http://phylopic.org>.

narratives found in the literature. The freshwater pearl mussel (*Margaritifera margaritifera*) shows relatively stable and neutral sentiment, reflecting its long standing threaten status and the existing challenges in measuring reintroduction outcomes. [Lavictoire and West, 2024]. The Eurasian beaver (*Castor fiber*) shows a clear upward sentiment trend, consistent with its successful reintroduction across Europe, having rebounded from near-extinction in the early 20th century [Needham et al., 2025]. In contrast, the European rabbit (*Oryctolagus cuniculus*) shows a declining trend, which aligns

with concerns expressed in the literature over population collapses in its native Iberian range due to habitat loss and disease. [Marín-García and Llobat, 2021].

Figure 8 shows overall scores by category which seem to increase over years, and with lower absolute value for more threatened categories. As Van Houtan et al. [2020] noted, species-level analysis tends to exaggerate positive tone for well-known, less threatened species, while the tone becomes more neutral or negative for critically endangered ones.

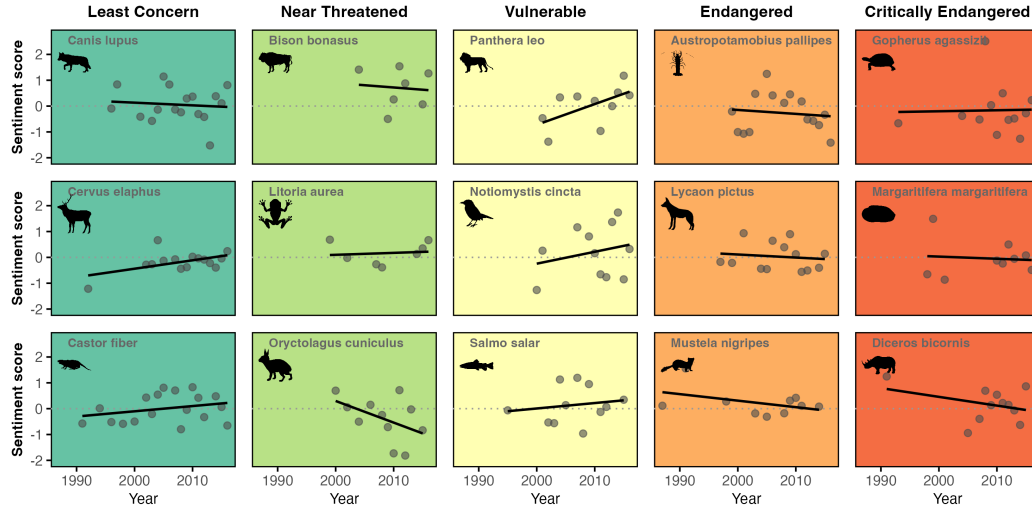


Figure 7: Sentiment scores for top 3 mentioned species for each threat category.

Note: Each column represents an IUCN category; three species per column are those with the highest mention counts. Silhouettes downloaded from <http://phylopic.org>.

There are also some visible outliers that could shift the direction of sentiment within a category. For example, in the *Diceros bicornis* box of Figure 7, dropping the early 90s data point would reverse the sign of the relationship. This underlines the importance of caution when interpreting sentiment means at species level, especially for sparsely studied taxa in some period.

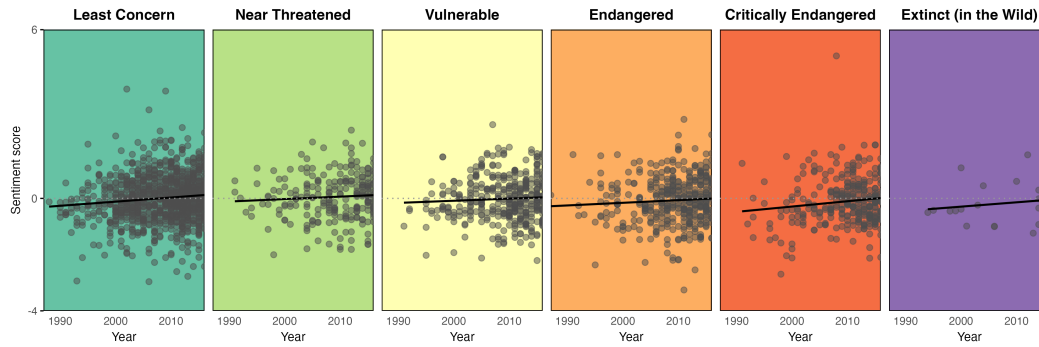


Figure 8: Sentiment scores by IUCN Red List category.

5.6 Spatial Patterns of Research Focus

The spatial distribution of reintroduction studies reveals clear global imbalances. As shown in Figure 9, most species–country mentions come from high-income, English-speaking regions such as the United States, Australia, and Western Europe. This pattern reflects broader geographic biases in conservation research, largely driven by the concentration of academic institutions in wealthier countries and the practical advantages of working locally, such as lower costs and fewer logistical barriers [Evans et al., 2022, Dos Santos et al., 2020]. This bias might also be extrapolated by our data

which contain only english, abstract as previously mentioned in the data limitations. Note, however, that the distribution of studies appears more balanced in more threatened categories.

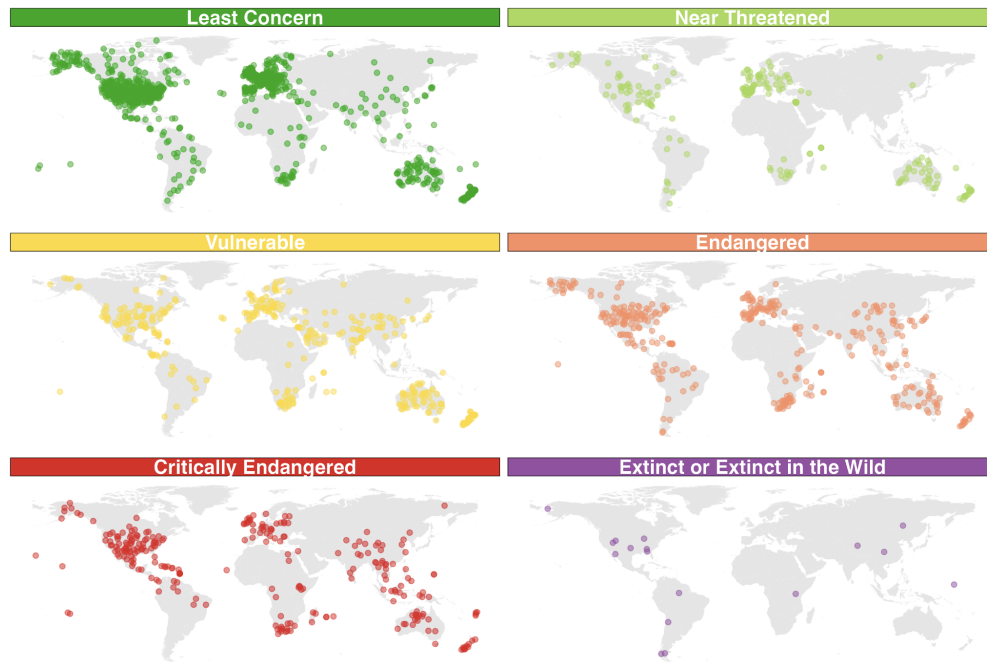


Figure 9: Geographical distribution of species–country mentions across abstracts (dot map).

Figure 10 then maps, for each country, the proportion of species studied that are categorised as threatened (Vulnerable, Endangered, Critically Endangered, or Extinct in the Wild). While northern countries dominate in terms of sheer volume, some southern regions—especially in Africa and Asia—show higher proportions of research focusing on threatened species. These findings echo those in Lennox et al. [2020], who argue that such spatial and taxonomic disconnects may distort global conservation priorities.

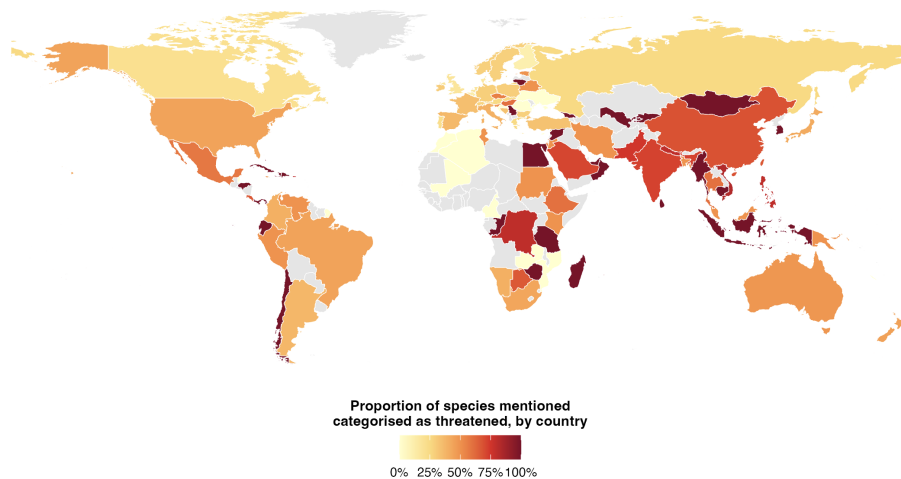


Figure 10: Proportion of species studied that are threatened, by country.

Note: Values are the share of mentions categorised as Vulnerable, Endangered, Critically Endangered, or Extinct in the Wild.

6 Discussion & Limitations

6.1 Conclusion

This study applied a transparent and reproducible pipeline to analyse over 4,000 abstracts in the species reintroduction literature. Several clear patterns emerged. First, sentiment scores have remained relatively stable over time, with a slight upward trend—although variability is high in earlier years due to smaller sample sizes. Second, sentiment varies by IUCN category and taxon: more positive tones are associated with less threatened and more charismatic species, while sentiment for critically endangered species tends to be more neutral or negative. Third, taxonomic and spatial biases remain strong. Mammals and birds dominate the literature, as do high-income countries, particularly English-speaking ones. These patterns reinforce what has already been observed in recent meta-analyses [Van Houtan et al., 2020, Lennox et al., 2020, Evans et al., 2022], but also reveal new insights at the intersection of sentiment and species focus.

Methodologically speaking, one of the main contributions of this study was the use of LLM-powered queries through external APIs. Ideally, one would compute a performance metric for each model or extraction method by manually labelling a representative sample of abstracts and comparing it against the automated output - but this was beyond the scope of the present study. While these models, like the one accessed here via Zhipu AI, offer promising opportunities for tasks such as geographic or even taxonomic extraction, they also come with important limitations. Most high-performance APIs are commercial and subject to prices or usage caps, which limits their scalability in larger studies (except if we put the price). From an environmental and ethical standpoint, they are also among the most energy-intensive NLP solutions.

For future projects, it may be more sustainable and robust to invest in improving well-established packages such as *taxize* and *rredlist*, which are open, interpretable, and adaptable. Enhancing these tools could help reduce reliance on commercial and tempting LLMs while promoting more frugal and reproducible research practices in biodiversity science.

6.2 Limitations

6.2.1 Geographical Bias. As discussed earlier, the spatial distribution of abstracts is heavily skewed toward high-income countries. This reflects broader publication and institutional biases in conservation science, but it also limits the generalisability of findings to tropical or low-income regions, which are often underrepresented despite their ecological importance.

6.2.2 Taxonomic Extraction Assumptions. Taxonomic mentions in abstracts are often multiple and not clearly prioritised. The method used in this study retained only the first valid binomial name extracted, assuming it to represent the focal species. While this holds true in most cases, manual checks revealed exceptions. Additionally, species mentioned by common name only (e.g., “elephant”) were not captured by the extraction algorithm, leading to incomplete data for those abstracts.

6.2.3 Temporal Mismatch in Red List Status. The IUCN threat status used in this study corresponds to present-day listings, while the abstracts cover nearly three decades (1987–2016). Some species may have shifted categories during this period, creating a mismatch between historical context and current classification. Moreover, reintroduction projects often target species that are regionally threatened but globally listed as “Least Concern”—a nuance that global Red List categories fail to capture.

6.2.4 Lexicon Bias in Sentiment Analysis. The sentiment lexicons applied here were developed for general-purpose contexts such as movie or product reviews. This creates a semantic mismatch when applied to conservation texts. As noted by Lennox et al. [2020], some words commonly used in ecological contexts—such as “wild” or “shark”—receive negative scores in standard lexicons, even though they should be neutral in reintroduction science. This misalignment likely introduces systematic bias into sentiment estimates.

6.2.5 Sample Attrition Bias. A significant number of abstracts were excluded at various stages of the pipeline due to missing or imprecise metadata. In many cases, the extraction process failed because abstracts did not mention a specific country, or referred only to broad regions (e.g., “Northern

America” or “Pacific”). Others were themselves meta-analyses or reviews without a clear species or geographic focus. These omissions reduce the overall coverage and may introduce selection bias.

6.3 Implications for Future Work

This study highlights both the potential and the limitations of using natural language processing to analyse trends in conservation science. Future work could build on this foundation in several ways.

First, developing or fine-tuning sentiment lexicons specifically tailored to ecological and conservation contexts would improve the reliability of sentiment scores. Domain-specific models trained on conservation literature—rather than general reviews—could help correct for semantic mismatches identified in this analysis.

Second, improving taxonomic and geographic entity recognition — particularly by enhancing the `taxize` package — could significantly improve species- and location-level inference. This includes more robust name resolution, better handling of taxonomic synonyms, and the ability to recognise common names alongside scientific binomials. Such improvements would make it easier to link species mentions with relevant ecological or regional priorities.

Third, addressing the temporal mismatch in species threat status would require modifications to the `rredlist` package, such as enabling access to historical IUCN Red List data. Incorporating time-specific threat categories or estimating species status at the date of publication would allow for more accurate assessments and stronger causal interpretations in time-based analyses.

More broadly, a reliable meta-analysis ultimately requires a database that is both exhaustive and representative. One natural next step would be to design a pipeline capable of collecting abstracts and metadata from multiple scientific repositories or online sources. Scraping large-scale content from the web could help overcome coverage gaps in existing datasets.

One limitation of most sentiment and NLP tools used in this study (and in the literature more generally) is that they are designed for English language text. In this context, one major advantage of LLM-based API querying is its multilingual capacity. With the right prompt, the same information can be extracted across languages in a harmonised way. While such approaches should be used cautiously and not as a default solution, they may prove valuable when no direct alternative exists.

References

- Darwin, C. (2013). *L'Origine des espèces: par le moyen de la sélection naturelle ou la préservation des races favorisées dans la lutte pour la vie*. Edited by Patrick Tort. Paris: Flammarion, Collection GF.
- Mayr, E. (1982). *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Harvard University Press.
- Woese, C. R., Kandler, O., & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, 87(12), 4576–4579.
- IUCN. (2012). *IUCN Red List Categories and Criteria: Version 3.1* (2nd ed.). Gland, Switzerland and Cambridge, UK: IUCN.
- Seddon, P. J., Griffiths, C. J., Soorae, P. S., & Armstrong, D. P. (2014). Reversing defaunation: Restoring species in a changing world. *Science*, 345(6195), 406–412.
- Lahire, B. (2023). *Les structures fondamentales des sociétés humaines*. Paris: La Découverte.
- Greenhalgh, T. (1997). Papers that summarise other papers: systematic reviews and meta-analyses. *BMJ*, 315(7109), 672–675.
- Van Houtan, K. S., Gagne, T., Jenkins, C. N., & Joppa, L. (2020). Sentiment analysis of conservation studies captures successes of species reintroductions. *Patterns*, 1(2), 100005. <https://doi.org/10.1016/j.patter.2020.100005>
- Evans, M. J., Gordon, I. J., Pierson, J. C., Neaves, L. E., Wilson, B. A., Brockett, B., Ross, C. E., et al. (2022). Reintroduction biology and the IUCN Red List: The dominance of species of Least Concern in the peer-reviewed literature. *Global Ecology and Conservation*, 38, e02242. <https://doi.org/10.1016/j.gecco.2022.e02242>
- Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6, 587–604.
- Lennox, R. J., Veríssimo, D., Twardek, W. M., Davis, C. R., & Jarić, I. (2020). Sentiment analysis as a measure of conservation culture in scientific literature. *Conservation Biology*, 34(2), 462–471. <https://doi.org/10.1111/cobi.13404>
- Pafilis, E., Frankild, S. P., Fanini, L., Faulwetter, S., Pavloudi, C., Vasileiadou, A., Arvanitidis, C., & Jensen, L. J. (2013). The SPECIES and ORGANISMS resources for fast and accurate identification of taxonomic names in text. *PLOS ONE*, 8(6), e65390. <https://doi.org/10.1371/journal.pone.0065390>
- Thessen, A. E., Cui, H., & Mozzherin, D. (2012). Applications of natural language processing in biodiversity science. *Advances in Bioinformatics*, 2012, 391574. <https://doi.org/10.1155/2012/391574>
- Lavictoire, L., & West, C. (2024). Population reinforcement of the endangered freshwater pearl mussel (*Margaritifera margaritifera*): Lessons learned. *Diversity*, 16(3), 187. <https://doi.org/10.3390/d16030187>
- Needham, R. J., Zabel, R. W., Roberts, D., & Kemp, P. S. (2025). The impact of reintroduced Eurasian beaver (*Castor fiber*) dams on the upstream movement of brown trout (*Salmo trutta*) in upland areas of Great Britain. *PLOS ONE*, 20(2), e0313648. <https://doi.org/10.1371/journal.pone.0313648>
- Marín-García, P. J., & Llobat, L. (2021). What are the keys to the adaptive success of European wild rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula? *Animals*, 11(8), 2453. <https://doi.org/10.3390/ani11082453>
- dos Santos, J. W., Correia, R. A., Malhado, A. C. M., Campos-Silva, J. V., Teles, D., Jepson, P., & Ladle, R. J. (2020). Drivers of taxonomic bias in conservation research: A global analysis of terrestrial mammals. *Animal Conservation*, 23(6), 679–688. <https://doi.org/10.1111/acv.12586>