International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I4P116

Volume 73 Issue 4, 159-179, April 2025


*Original Article*

# Speech Translation: A Bibliometric Analysis of Research Trends and Contributions based on Scopus Data (2000-2024)

Maria Labied[1], Abdessamad Belangour[2], Mouad Banane[3]

[1,2]*Laboratory of Information Technology and Modeling, LTIM Hassan II University, Ben M'sik Faculty of Sciences Casablanca, Morocco.*
[3]*Department, Laboratory of Artificial Intelligence & Complex Systems Engineering, Hassan II University, Faculty of Legal, Economic, and Social Sciences Casablanca, Morocco.*

[1]*Corresponding Author : mr.labied@gmail.com*

Received: 2 December 2024 Revised: 8 March 2025 Accepted: 12 March 2025 Published: 26 April 2025

***Abstract*** *- This study provides a comprehensive bibliometric analysis of speech translation research from 2000 to 2024, leveraging Scopus database data to identify key trends, influential contributions, and collaborative networks in this rapidly evolving field. We map the transition from traditional statistical methods to advanced neural and deep learning approaches in speech translation technologies by analysing publication patterns, citation metrics, and research themes. Our findings highlight the most prolific authors, institutions, and countries, along with the leading journals and conferences that serve as primary outlets for high-impact research. Notably, the analysis reveals a substantial increase in research activity and a growing focus on end-to-end translation systems and multilingual corpora, demonstrating the field's shift towards scalable and effective real-world applications. The importance of international collaborations and interdisciplinary research is emphasized, showcasing their role in driving innovation and addressing complex challenges. This bibliometric analysis provides valuable insights for researchers, practitioners, and policymakers, offering a foundational understanding of the current landscape and future directions of speech translation research. By elucidating the dynamics of this field, our work aims to inspire further advancements and enhance the impact of future research efforts.*

***Keywords*** *- Speech Translation, Bibliometric Analysis, End-to-End speech translation, Direct speech translation, Machine translation, Automatic speech translation.*

## 1. Introduction

Speech translation technology has become an integral part of modern communication, bridging language barriers and facilitating global interactions. The development of this technology dates back to the mid-20th century when early attempts focused on Automatic Speech Recognition (ASR) and Machine Translation (MT) as separate fields. The integration of these technologies has led to the creation of speech-to-text translation systems, which transcribe spoken language into text in the source language and then translate it into the target language. The evolution of speech translation technology can be traced through significant milestones. In the 1980s, advances in digital signal processing and statistical methods enabled the development of more accurate ASR systems. By the 1990s, Statistical Machine Translation (SMT) techniques had improved[1], leading to the prototype speech translation systems. These early systems demonstrated the feasibility of real-time speech translation, albeit with limited accuracy and vocabulary. The 21st century has seen exponential growth in this field, driven by advancements in deep learning and neural networks. Neural Machine Translation (NMT) has replaced traditional SMT, offering more fluent and contextually accurate translations. Google's introduction of neural networks in its Translate service in 2016 marked a significant leap forward, improving both the quality and speed of translations [2]. A significant advancement in speech translation technology is the development of end-to-end models. Unlike traditional cascade models that separate ASR and MT processes, end-to-end models directly convert speech in the source language into text in the target language without an intermediate text representation[3-5]. This approach simplifies the translation pipeline, reduces latency, and minimizes error propagation between ASR and MT components. End-to-end models, such as those based on sequence-to-sequence architectures with attention mechanisms, have demonstrated remarkable improvements in translation quality and efficiency[6-8]. The impact of end-to-end models is particularly evident in real-time applications,

where reducing latency is critical. These models enable faster and more accurate translations, enhancing user experience in live conversations and broadcasts. Additionally, end-to-end models can be more easily adapted to low-resource languages, promoting linguistic diversity and inclusivity in digital communication[9, 10]. As the technology continues to advance, researchers are focusing on improving accuracy, reducing latency, and handling a broader range of languages and dialects[11-14]. Integrating context-aware and culturally sensitive translation models is a growing area of interest, aiming to provide more nuanced and appropriate translations. Developing end-to-end speech translation systems, which bypass the intermediate text representation, represents a promising direction for future research [15].

Speech translation has emerged as a critical technology for overcoming language barriers in an increasingly interconnected world. The field has evolved from early rule-based and statistical machine translation approaches to modern neural-based models that leverage deep learning and self-supervised techniques. Given the rapid advancements in artificial intelligence, speech processing, and multilingual model architectures, understanding the trajectory of research in this domain is essential for identifying key developments and future directions. Examining trends in speech translation research provides valuable insights into how methodologies have evolved, which areas are receiving the most attention, and what challenges remain. A bibliometric analysis enables a quantitative evaluation of publication patterns, citation impact, and research collaborations, helping to identify influential contributions, emerging research themes, and shifts in technological focus. Such an analysis is particularly significant in highlighting the transition from statistical machine translation to end-to-end neural systems and the increasing role of multilingual corpora, low-resource language translation, and real-time applications.

Furthermore, understanding the research landscape in speech translation is crucial for addressing current limitations and guiding future innovations. As models become more sophisticated, issues such as bias in translation, ethical concerns in data collection, and the computational cost of large-scale models continue to shape the field. By identifying key trends, this study contributes to a deeper comprehension of how speech translation technologies are evolving and how they can be improved to enhance multilingual communication on a global scale. The primary purpose of this bibliometric analysis is to provide a comprehensive overview of the research landscape in the field of speech translation. By analyzing key trends, influential contributions, and collaborative networks, this study aims to identify the most significant research directions, methodologies, and technological advancements that have shaped the development of speech translation systems. This includes both speech-to-speech (S2ST) and speech-to-text (STT) translation technologies. The scope of this bibliometric analysis encompasses a detailed examination of publication trends over time to understand the growth and maturation of speech translation research. It identifies the most prolific authors, institutions, and countries contributing to the field. The study also includes an analysis of citation patterns to highlight the most influential papers and their impact on subsequent research.

Furthermore, it explores research themes and topics through keyword co-occurrence and thematic mapping. Another important aspect is the assessment of collaborative networks and patterns of international cooperation among researchers and institutions. Finally, it evaluates the leading journals and conferences that serve as primary publication outlets for speech translation research. Several research questions guide this study. The first question examines how the volume of research publications in the field of speech translation has evolved. It hypothesizes that there has been a significant increase in the number of publications on speech translation, reflecting growing academic and industry interest. The second question identifies the most prolific authors and which institutions and countries have substantially contributed to speech translation research. It hypothesizes that a small number of key researchers and leading institutions dominate the field, with notable contributions from technologically advanced countries.

The third question investigates which papers are the most highly cited and their impact on developing speech translation technologies. The hypothesis is that certain seminal papers have disproportionately influenced the field, serving as foundational works for subsequent research. The fourth question explores the main research themes and topics within speech translation and how these themes have evolved. The hypothesis here is that research themes have shifted from foundational techniques to advanced deep learning methods and real-world applications, indicating a maturing field.

The fifth question analyzes the collaboration patterns among researchers and institutions and how these networks facilitate the advancement of speech translation technology. The hypothesis suggests that collaborative efforts, particularly international collaborations, are crucial in driving innovation and addressing complex challenges in speech translation. The sixth question examines which journals and conferences are the primary platforms for disseminating speech translation research and their relative impact. The hypothesis posits that a few specialized journals and conferences have emerged as the leading platforms for publishing high-impact research in the field. To answer these questions, this study employs a comprehensive bibliometric approach. Data is collected from major scientific databases such as Scopus, using relevant keywords and search queries specific to speech translation research. The study applies various bibliometric tools and techniques, including citation analysis, co-authorship network analysis, and keyword co-occurrence analysis, to extract

meaningful insights from the collected data. Publication trends are examined by analyzing the number of publications over time, while the most prolific authors, institutions, and countries are identified based on publication counts and citation metrics. Highly cited papers are highlighted to understand their impact on the field.

Research themes are mapped through keyword analysis, revealing the evolution of topics over time. Collaborative networks are assessed by analyzing co-authorship patterns, and the leading journals and conferences are evaluated based on their publication and citation records. This systematic approach ensures a thorough and objective analysis, providing a detailed and insightful understanding of the speech translation research landscape.

## 2. Methodology

### *2.1. Data Collection*

The primary data source for this bibliometric analysis is Scopus, one of the largest and most comprehensive abstract and citation databases in the peer-reviewed literature. Scopus offers extensive coverage of scientific journals, conference proceedings, and other scholarly publications, making it an ideal resource for conducting a thorough bibliometric study. The database provides detailed citation information and allows for advanced search queries, enabling the extraction of relevant data to analyse speech translation research.

A series of carefully selected keywords and search queries were employed to ensure a comprehensive and focused collection of relevant literature. These keywords were chosen based on their relevance to the field of speech translation and their ability to capture a broad range of related studies. The keywords and search queries used include:

- Speech Translation: This primary keyword was used to identify general studies and publications related to the field of speech translation.
- Speech-to-Speech Translation (S2ST): This keyword was used to specifically target research focused on directly translating spoken language into another spoken language.
- Speech-to-Text Translation (STT): This keyword targeted studies and publications that involve translating spoken language into written text in a different language.
- Text-to-Speech Synthesis (TTS): This term was used to find studies on the synthesis of speech from translated text, completing the speech-to-speech translation process.
- Multilingual Speech Processing: This keyword was used to capture research that deals with processing and translating speech across multiple languages.
- Speech-to-Text Synthesis: This term was used to find studies that explore the synthesis of written text from spoken language translations.
- Automatic Speech Recognition (ASR) AND Machine Translation (MT): This combined keyword search was used to capture studies that discuss the integration of speech recognition with machine translation, a critical component of many speech translation systems.
- Machine Translation Metrics: This keyword was included to identify studies that evaluate speech translation systems using various machine translation metrics.
- BLEU Score: This specific keyword was used to capture research that employs the BLEU score metric for evaluating the quality of translations produced by speech translation systems.

The search queries were structured to include these keywords in various combinations and were adjusted to account for different spellings and terminology used in the literature. Boolean operators (AND, OR) were employed to refine the searches and ensure comprehensive coverage of relevant studies.

Additionally, filters were applied to limit the search results to peer-reviewed articles, conference papers, and review articles to ensure the quality and relevance of the included studies. The search timeframe was set to cover publications from 2000 to the most recent advancements in speech translation technology, providing a comprehensive overview of the field's evolution.

A robust dataset of relevant publications was compiled for the bibliometric analysis by employing these keywords and search queries in Scopus. This dataset serves as the foundation for subsequent analyses, including publication trends, citation patterns, research themes, and collaborative networks, offering valuable insights into the state and development of speech translation research. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed to ensure a systematic and transparent study selection process. The PRISMA flow diagram was used to document each stage of the study selection process, from initial identification to the final inclusion of studies.

This approach enhances the reproducibility and transparency of the bibliometric analysis. The flow diagram consists of four main phases: identification, screening, eligibility, and inclusion. Records were retrieved from the Scopus database using the specified keywords and search queries during identification.

In the screening phase, duplicate records were removed, and the remaining records were screened based on their titles and abstracts. The eligibility phase involved thoroughly assessing the full-text articles' relevance and suitability for inclusion. Finally, the inclusion phase documented the studies that met all criteria and were included in the bibliometric analysis (Figure 1).

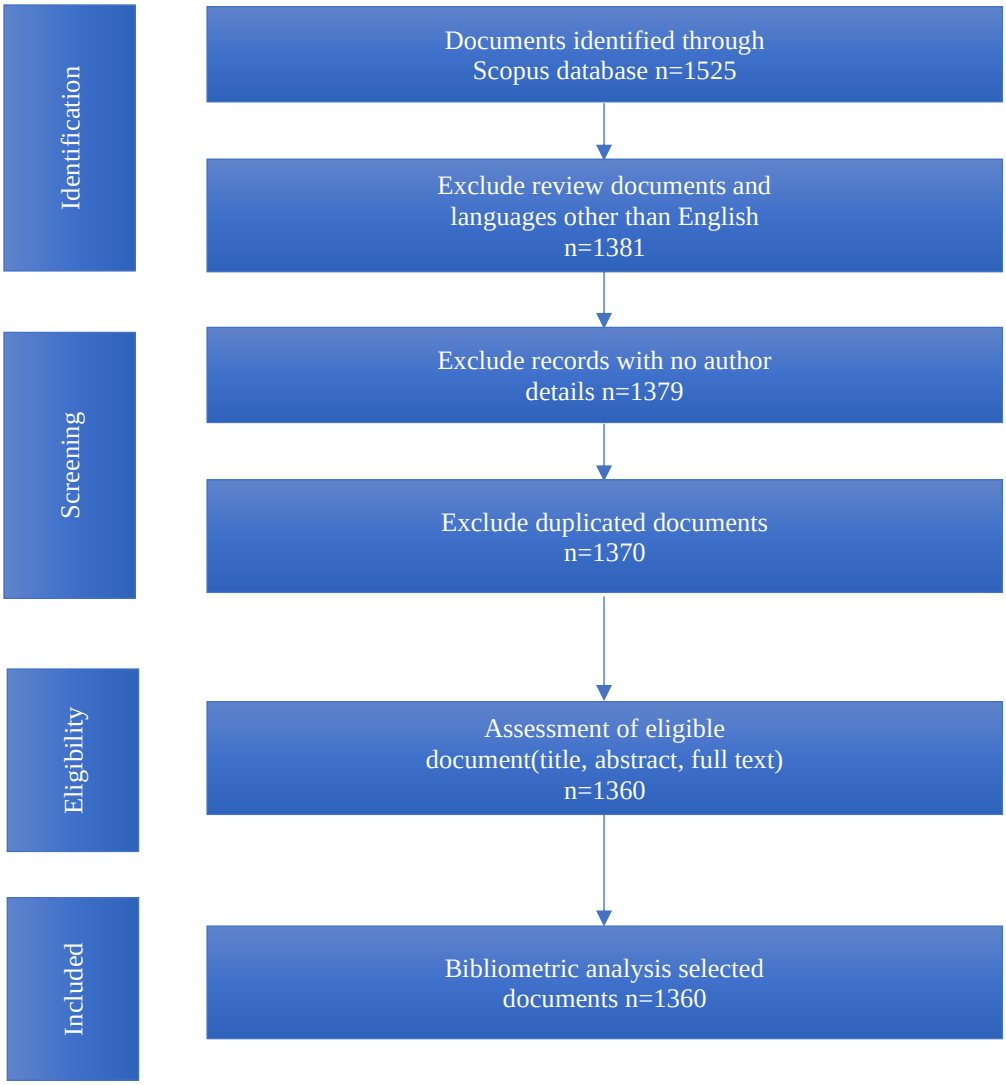


**Fig. 1 PRISMA flow diagram**

### *2.2. Inclusion and Exclusion Criteria*

An inclusion and exclusion criteria was established to ensure the relevance and quality of the studies included in this bibliometric analysis. The criteria were designed to filter out studies that do not directly contribute to the field of speech translation or do not meet the necessary quality standards. The inclusion criteria are as follows:

- Relevance to Speech Translation: Only studies explicitly focused on speech translation, including both speech-to-speech (S2ST) and speech-to-text (STT) translation technologies, were included. This includes research on automatic speech recognition (ASR), machine translation (MT), and text-to-speech (TTS) synthesis as they relate to speech translation systems.
- Peer-Reviewed Publications: To ensure the credibility and scientific rigor of the included studies, only peer-reviewed journal articles and conference papers were considered.
- Language: Publications must be in English to ensure accessibility and comprehensibility for the analysis.
- Accessibility: Full-text articles must be accessible through the Scopus database or other academic sources to allow for thorough review and analysis.

The exclusion criteria are as follows:

- Irrelevant Studies: Articles that do not primarily focus on speech translation or are tangentially related, such as those focusing solely on speech recognition or machine translation without application to speech translation, were excluded.
- Non-Peer-Reviewed Publications: Books, editorials, opinion pieces, and non-peer-reviewed conference abstracts were excluded to maintain the quality and reliability of the dataset.

- Duplicate Publications: Duplicate records identified during the data collection process were removed.

- Non-English Publications: Articles unavailable in English were excluded due to language constraints.

### *2.3. Timeframe of the Study*

The timeframe for this bibliometric analysis was set to encompass the evolution of speech translation research from its early foundational works to the most recent advancements. The selected timeframe spans from 2000 to 2024. This period was chosen for several significant reasons. The year 2000 marks the beginning of significant technological advancements in speech translation, including the development and implementation of machine learning and deep learning techniques. These advancements have dramatically transformed the field, leading to substantial improvements in accuracy and functionality. Also, the research publications and citation databases have become more comprehensive and consistent in their coverage from 2000 onwards. This ensures a more reliable and robust dataset for bibliometric analysis, as earlier records may be incomplete or inconsistently indexed.

Additionally, analyzing research from 2000 to 2024 allows for focusing on recent trends, innovations, and the current state of the art in speech translation. This period captures the most impactful and relevant developments that are shaping the present and future of the field. By applying these inclusion and exclusion criteria within the specified timeframe, the study ensures a focused and high-quality dataset for subsequent bibliometric analysis.

### *2.4. Bibliometric Tools and Techniques*

Several advanced software tools and techniques were utilized to conduct a thorough bibliometric analysis of speech translation research. These tools are well-suited for handling large datasets and performing complex analyses, ensuring a comprehensive and detailed examination of the research landscape. The primary tools used include VOSviewer, which facilitates the visualization of co-authorship, co-citation, and keyword co-occurrence networks. VOSviewer is particularly useful for identifying research trends and patterns in large datasets, making it an invaluable tool for this study.

CiteSpace is used for detecting and visualizing emerging trends and transient patterns in citation networks, making it an excellent choice for identifying key areas of innovation and development in speech translation research. CiteSpace's ability to highlight critical turning points in research and visualize the evolution of scientific fields enhances the depth of the bibliometric analysis. R is used for citation analysis, thematic mapping, and collaboration network analysis, offering flexibility and robustness in data analysis. The integration of R allows for custom analyses and the creation of detailed visualizations tailored to the specific needs of this study. Scopus Analysis Tools offer quick insights into publication trends, citation analysis, and author collaboration. These tools were used for preliminary analysis and to validate findings from other software tools. By utilizing these advanced software tools, the study ensures a comprehensive and detailed analysis of the speech translation research landscape. The combination of VOSviewer, R, and Scopus analysis tools provides a robust and multi-faceted approach to bibliometric analysis, enabling the identification of key trends, influential contributions, and collaborative networks within the field. This methodological approach enhances the reliability and depth of the findings, contributing valuable insights to the ongoing development of speech translation technologies.

### *2.5. Analysis Metrics*

Several key bibliometric metrics were analysed to gain a comprehensive understanding of the research landscape in speech translation. These metrics provide insights into the volume, impact, and collaborative nature of research in this field. The primary metrics analyzed include:

- Citation Count: This is a fundamental metric that indicates the impact and recognition of a research paper within the scientific community. By analyzing citation counts, the study identifies the most influential papers and authors in speech translation research.
- Analyzing Publication Trends Over Time: reveals the growth and development of speech translation research. This metric helps understand how the field has evolved and identifies significant research activity and innovation periods.
- Co-Authorship Networks Analysis: examines the collaborative relationships between researchers. This metric helps identify key research groups, institutions, and countries contributing to speech translation research and highlights patterns of international collaboration.
- Keyword Co-Occurrence Analysis: identifies the main research themes and topics within speech translation. The study uncovers emerging trends and shifts in research focus over time by examining the frequency and relationships between keywords.

This metric helps identify influential papers and how they are interconnected within the research landscape, providing insights into the intellectual structure of the field. Thematic mapping involves clustering related research topics and visualizing their relationships. This metric helps understand the broader themes and subfields within speech translation research, facilitating a holistic view of the field's development.

## 3. Results and Discussion

### *3.1. General Publication Trends*

#### *3.1.1. Yearly Publication Trends*

This analysis highlights the historical trends in publications related to speech translation within the Scopus database. The

early years saw gradual growth, with a noticeable increase in 2006. From 2007 to 2011, annual publications fluctuated modestly. However, significant growth began in 2012, peaking dramatically in 2023 with nearly 250 publications. This surge indicates a heightened interest and rapid advancements in speech translation technologies, as illustrated in Figure 2. The cumulative trend underscores a consistent rise in research activity, highlighting the field's increasing importance and suggesting ongoing technological and methodological advancements.

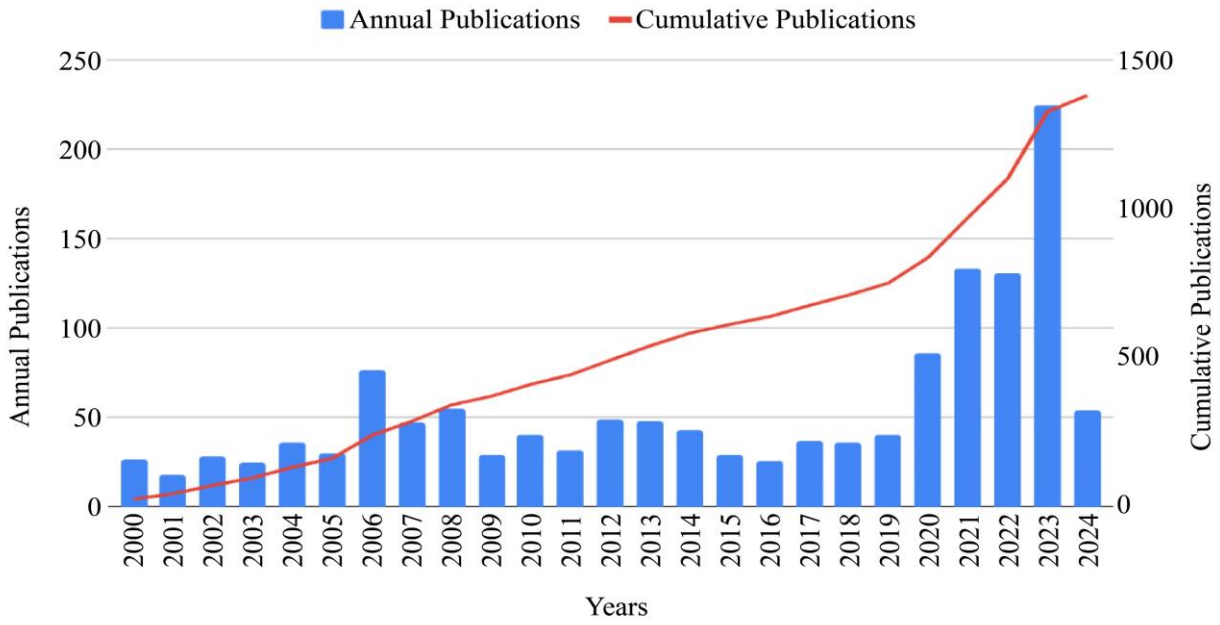


**Fig. 2 Speech translation yearly trend**

*3.1.2 Most Prolific Authors, Countries*

Based on the bibliometric data, the leading author in the field of speech translation is Nakamura, S., with 71 publications, making a substantial impact on the development and advancement of speech translation technologies. This high number of publications indicates a strong commitment to research and possibly a leadership role in multiple projects or research groups.

Nakamura's contributions likely cover various aspects of speech translation, including algorithm development, system integration, and practical applications. Following closely is Waibel, A., with 64 documents. Waibel's prolific output suggests a deep involvement in the field, possibly through both academic research and applied projects.

This level of productivity could be due to collaborations with different institutions or leading significant projects that push the boundaries of speech translation capabilities. The impact of Waibel's work is likely broad, influencing both theoretical advancements and practical implementations. Negri, M. and Turchi, M. follow with 44 and 43 documents, respectively.

Their close publication counts indicate they might work on similar projects or within the same research network. Their contributions are crucial for advancing specific areas within speech translation, such as improving accuracy, dealing with multilingual contexts, or integrating speech translation with other AI technologies. Their work is essential for overcoming current limitations and pushing the field forward. The list also includes other notable contributors like Sakti, S., Watanabe, S., and Pino, J., among others, each with significant publication counts ranging from 30 to 40. These authors represent a strong core of researchers who drive innovation and development in speech translation. Their collective work likely covers various topics, from fundamental linguistics and machine learning research to developing robust, real-world applications. The diversity in their research areas and methodologies contributes to the field's growth and evolution.

The chart of the 10 most prolific authors in speech translation (Figure 3) highlights the significant contributions of a dedicated group of researchers. Their work spans various aspects of the field, driving both theoretical advancements and practical applications. The high publication counts reflect their active engagement in pushing the boundaries of what is possible in speech translation, paving the way for future more sophisticated and accurate translation systems. The contributions of these authors also underscore the collaborative nature of the research community. Many of these prolific authors are likely involved in international collaborations, contributing to a global effort to enhance speech translation technology. This collective effort is crucial in addressing the

complexities of translating speech across different languages and dialects, ensuring the technology can be applied universally. Furthermore, the presence of multiple authors with high publication counts indicates a competitive yet productive research environment. This competition drives innovation as researchers strive to publish high-quality work that addresses current challenges in the field. It also suggests that funding and resources are being effectively allocated to support extensive research in speech translation, enabling sustained progress.

The speech translation field has seen significant contributions from various countries worldwide. The analysis of the most prolific countries in this domain reflects the global nature of research and development efforts. Each country's contributions are marked by the number of documents produced, indicating their active engagement and impact in advancing speech translation technologies (Figure 4). The United States leads with 419 publications, demonstrating its prominent role in the field. This leadership position is likely supported by substantial investments in research and development, a robust academic infrastructure, and collaborations between universities, research institutions, and industry.

The high volume of publications from the United States signifies a diverse range of research activities, from theoretical studies to practical applications, driving innovation and setting standards in speech translation. Japan follows with 229 publications, highlighting its significant contributions to speech translation research. Japan's strong emphasis on technology and innovation, supported by both governmental and industrial initiatives, has fostered a conducive environment for advancements in this field.

The focus on speech translation in Japan is also driven by its practical applications in addressing language barriers within a multilingual society and enhancing international communication. China, with 178 publications, is another major contributor to the field. The rapid growth of technological research and development in China, coupled with substantial governmental support, has propelled its prominence in speech translation. Chinese researchers have been actively exploring various aspects of speech translation, including machine learning algorithms, natural language processing, and real-time translation systems, contributing to the global body of knowledge and technological advancements. Germany and India also feature prominently, with 147 and 91 publications, respectively. Germany's strong engineering and technical research capabilities, supported by renowned universities and research institutions, have made significant strides in speech translation.

India, on the other hand, has seen a growing interest in this field driven by its diverse linguistic landscape and the need for effective communication tools. Both countries are making important contributions to advancing speech translation technologies, addressing unique challenges, and exploring innovative solutions. Other notable contributors include France, Spain, Italy, and the United Kingdom, each producing a substantial number of publications. These countries have established research communities and collaborative networks that drive forward the development and implementation of speech translation technologies. The cumulative efforts of these and other contributing countries illustrate the global nature of speech translation research, characterized by shared knowledge, collaborative initiatives, and a common goal of breaking down language barriers through advanced technology.

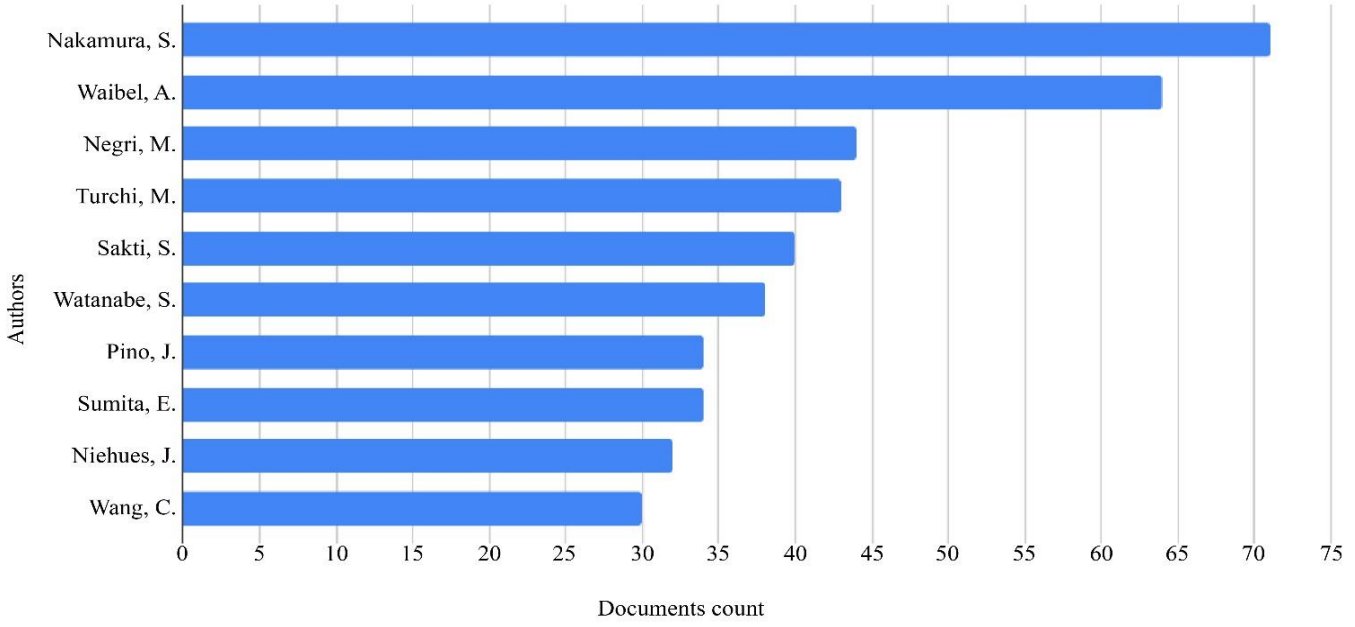


**Fig. 3 10 Most prolific authors in the field of speech translation**

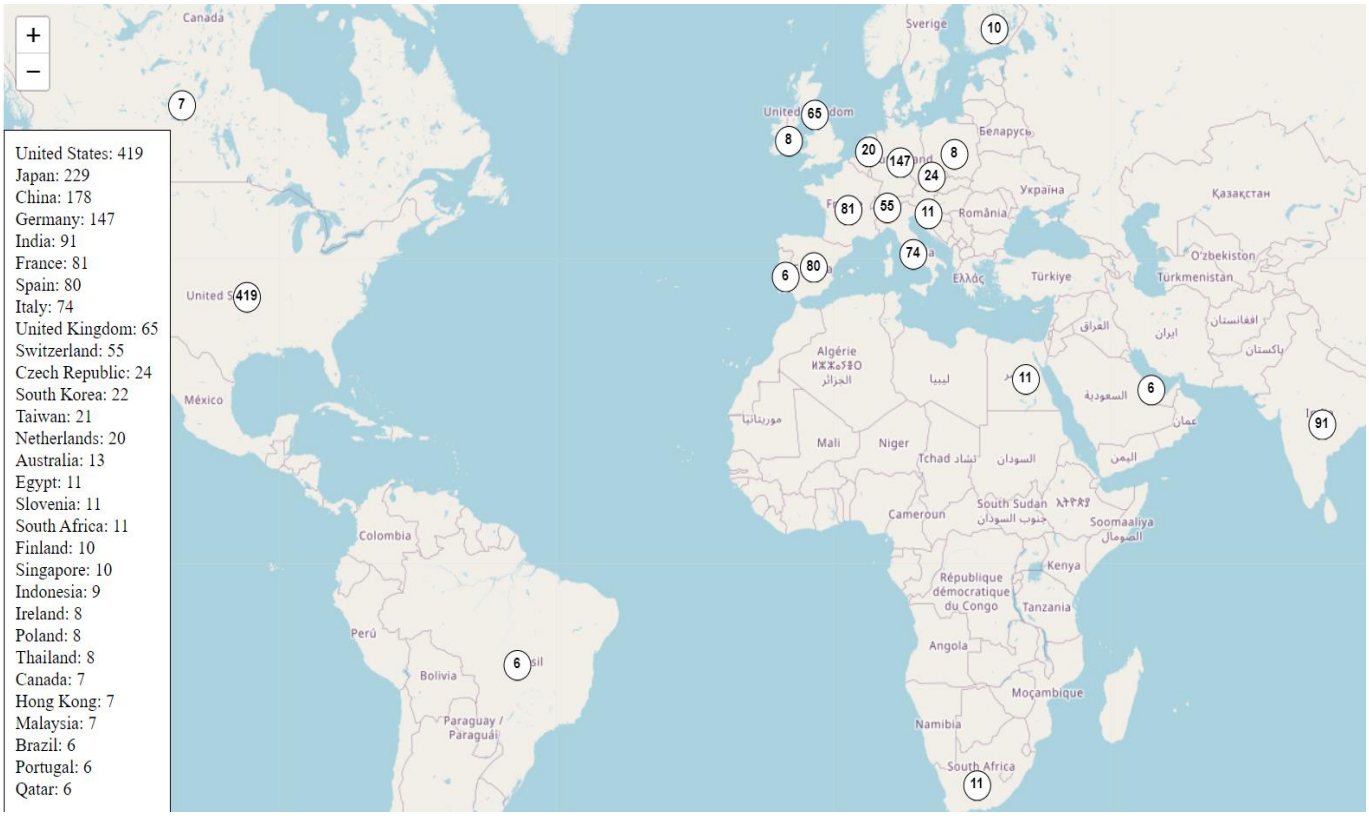


**Fig. 4 Most prolific countries in speech translation**

### *3.2. Most Cited Papers and Their Impact*

The analysis of the most cited papers and their impact highlights influential works within the field of speech translation, showcasing their significance based on citation counts (Table I). The paper "Sign-to-speech translation using machine-learning algorithms," published in 2020 and cited 549 times, has significantly advanced the application of machine learning in speech translation, demonstrating high relevance and utility in the field. The 2019 study "A Comparative Study on Transformer vs RNN in Speech Translation," with 458 citations, underscores the importance of model comparison and optimization in improving translation accuracy. "MUST-C: A multilingual speech translation corpus" from 2019, cited 263 times, has provided a critical resource for multilingual translation research, facilitating diverse language applications and further studies. The 2005 "Edinburgh System Description for the 2005 IWSLT Evaluation Campaign" (258 citations) and the 2002 "Phrase-based statistical machine translation" (212 citations) have contributed foundational methodologies, influencing subsequent developments in statistical and system-based translation approaches. The 2017 paper "Sequence-to-sequence models can directly translate foreign speech" (203 citations) highlighted the potential of end-to-end models, while "Making machines understand us in reverberant rooms" (2012, 201 citations) addressed challenges in speech recognition accuracy under varying acoustic conditions. Earlier works like "Toward a broad-coverage Bilingual Corpus for Speech Translation" (2002, 177 citations) and "Computing Consensus Translation from multiple machine translation systems" (2006, 148 citations) have laid important groundwork in corpus development and consensus translation techniques. Finally, "Recent developments on ESPNeT toolkit boosted the performance of speech translation tasks" (2021, 164 citations) showcased advancements in toolkit performance, boosting translation task efficiency. These papers collectively illustrate the dynamic progress and critical innovations in speech translation, significantly shaping the field through their high impact and continued relevance in contemporary research.

**Table 1. 10 Most cited papers in the field of speech translation**

| # | Title | Year | Citations | Source Title |
|---|---|---|---|---|
| 1 | Sign-to-speech translation using machine-learning algorithms-assisted stretchable sensor arrays[16] | 2020 | 549 | Nature Electronics |
| 2 | A Comparative Study on Transformer vs RNN in Speech Translation[17] | 2019 | 458 | 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) |

| | | | | |
|---|---|---|---|---|
| 3 | MUST-C: A multilingual speech translation corpus[18] | 2019 | 263 | NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies |
| 4 | Edinburgh System Description for the 2005 IWSLT Evaluation Campaign[19] | 2005 | 258 | 2005 International Workshop on Spoken Language Translation (IWSLT 2005) |
| 5 | Phrase-based statistical machine translation[20] | 2002 | 212 | Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) |
| 6 | Sequence-to-sequence models can directly translate foreign speech[21] | 2017 | 203 | Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH |
| 7 | Making machines understand us in reverberant rooms: robustness against reverberation for automatic speech recognition and speaker identification[22] | 2012 | 201 | IEEE Signal Processing Magazine |
| 8 | Toward a broad-coverage bilingual corpus for speech translation[23] | 2002 | 177 | Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002) |
| 9 | Recent developments in the ESPNeT toolkit boosted the performance of speech translation tasks[24] | 2021 | 164 | ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings |
| 10 | Computing consensus translation from multiple machine translation systems[25] | 2006 | 148 | EACL 2006 - 11th Conference of the European Chapter of the Association for Computational Linguistics |

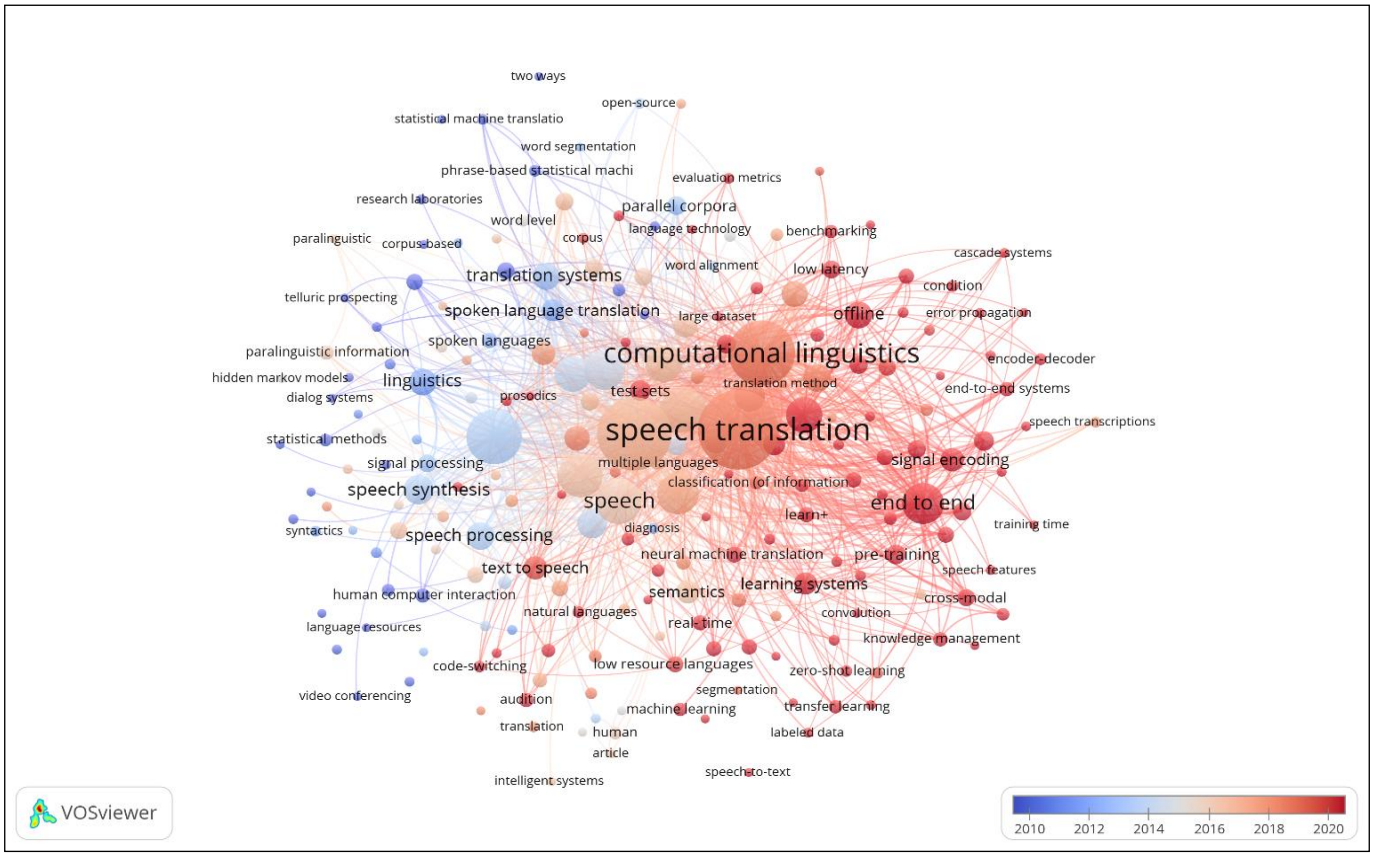


**Fig. 5 Knowledge map overlay visualization of speech translation based on keywords co-occurrence**

### *3.3. Research Themes and Topics*

*3.3.1. Keyword Co-Occurrence Analysis*

The field of speech translation is characterized by a dynamic and interdisciplinary research environment, as evidenced by a comprehensive keyword co-occurrence analysis (Figure 5). At the core of this analysis are the terms "speech translation" and "computational linguistics," which dominate the central theme, reflecting their pivotal role in the domain. The dense network of related keywords surrounding these central terms indicates the extensive integration of linguistic theories and computational methods essential for advancing speech translation technologies. Prominent research areas within this field include speech processing and speech synthesis. Keywords such as "speech processing," "speech synthesis," and "speech recognition" suggest a strong focus on the foundational technologies necessary for effective speech translation. This area of research explores the conversion of spoken language into text and vice versa, which is critical for developing accurate and efficient translation systems. Another significant cluster centers on the development of end-to-end systems, highlighted by keywords like "end-to-end," "neural networks," and "deep learning." This reflects the growing interest in creating end-to-end speech translation systems that utilize advanced machine learning techniques. These systems aim to directly translate spoken language without the need for intermediate text representations, thereby enhancing both speed and accuracy. Research efforts are also heavily focused on integrating linguistic knowledge into translation systems, as indicated by keywords like "translation systems," "spoken language translation," and "linguistics." This research aims to improve the accuracy and fluency of translations by incorporating syntactic and semantic information, addressing the inherent complexities of language structure. The map also reveals a robust research focus on developing and refining computational models and algorithms, with keywords such as "machine translation," "statistical models," and "neural machine translation" being prominent. These models are crucial for the translation process, and advancements in this area contribute to more reliable and nuanced translations.Emerging trends in the field are evident through the presence of newer keywords such as "zero-shot translation," "multimodal translation," and "semantic learning." These terms indicate innovative approaches being explored within the research community. For example, zero-shot translation involves translating between language pairs without direct training data, while multimodal translation integrates additional data sources, such as visual context, to improve translation quality. The interconnected nature of the keywords on the map underscores the collaborative and interdisciplinary efforts within the speech translation research community. The overlaps between clusters suggest that advancements in one area, such as deep learning, often benefit related areas like speech processing and translation systems, driving overall progress in developing sophisticated speech translation technologies. Additionally, the knowledge map density visualization provides further insights into the intensity and focus of research activities (Figure 6).

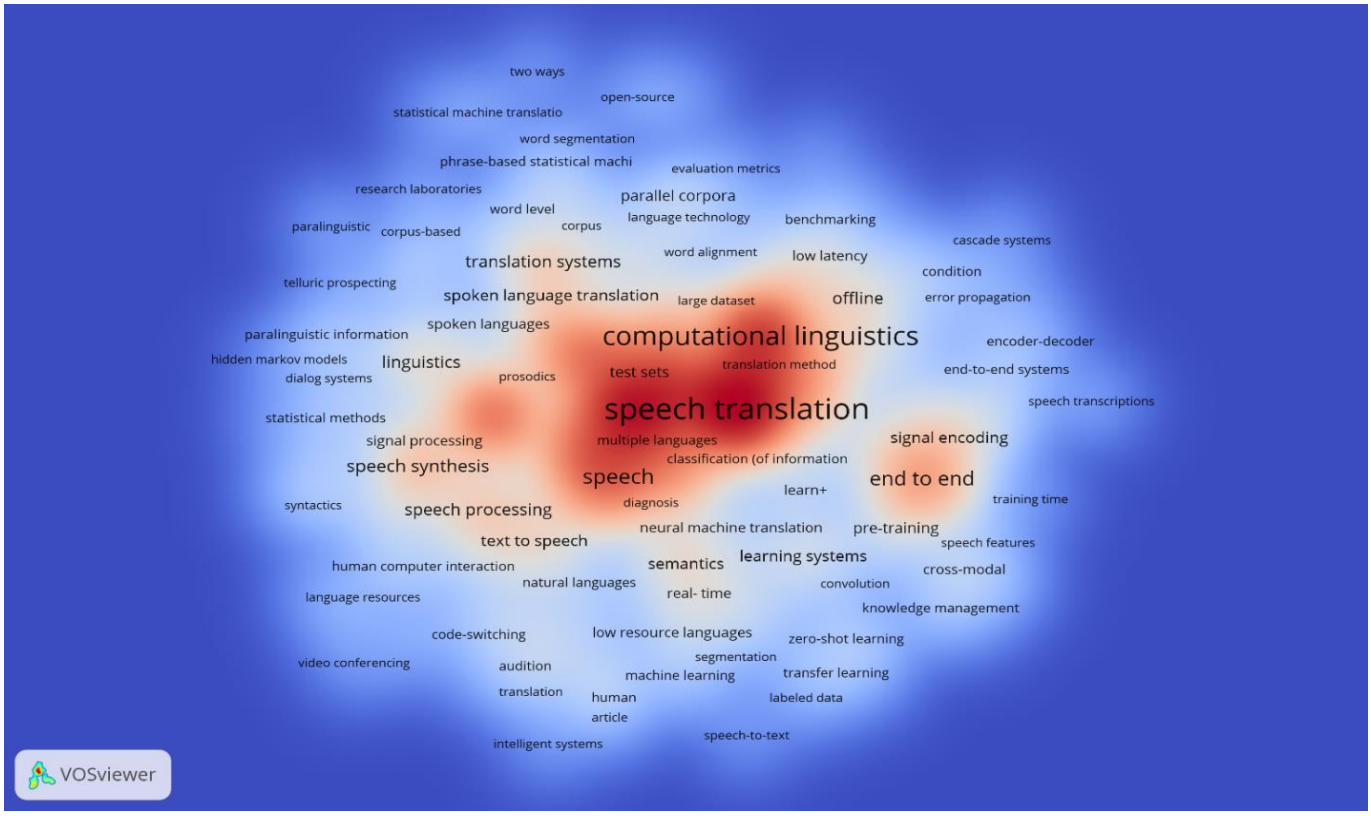


**Fig. 6 Knowledge map density visualization of speech translation based on keywords co-occurrence**

The heatmap-style visualization highlights areas with higher keyword occurrences, indicating more active research and development. Central terms like "speech translation," "computational linguistics," "speech processing," and "end-to-end" are shown in warmer colors, emphasizing their prominence and the significant attention they receive from researchers. This visualization helps identify key areas of interest and emerging hotspots within the field, guiding future research directions and resource allocation.

*3.3.2. Emerging and Declining Research Themes*

Emerging themes such as end-to-end systems, neural networks, and deep learning are becoming increasingly prominent. These approaches aim to create seamless translation systems that enhance accuracy and efficiency by bypassing intermediate text representations. Additionally, innovative methods like zero-shot translation and multimodal translation are gaining traction, reflecting the research community's effort to address complex translation tasks and integrate additional data sources for improved quality. Conversely, traditional approaches such as statistical machine translation (SMT) and rule-based translation are witnessing a decline. The reduced number of publications in recent years for these themes indicates a shift towards more modern, data-driven methods.

The rise of Neural Machine Translation (NMT) and advanced machine learning techniques has overshadowed these older techniques, leading to a decreased focus on purely linguistic rule-based systems and offline translation solutions. This transition highlights the research community's preference for more efficient and accurate translation technologies that handle real-time and online processing. Table II illustrates the number of publications per year for various emerging and declining themes in speech translation research. This data provides a clear picture of how the focus has shifted over time, highlighting the growing interest in advanced machine learning techniques and the decline of older, less efficient methods (Figure7).

Recent advancements in speech translation research have shifted towards unsupervised and self-supervised learning techniques, addressing key challenges such as data scarcity and domain adaptation. Traditionally, speech translation models have relied heavily on supervised learning, which requires large amounts of manually labeled parallel data. However, obtaining such datasets, especially for low-resource languages, remains a significant challenge. Unsupervised learning offers a promising alternative by leveraging unannotated speech and text data to improve translation performance without explicit supervision. Integrating unsupervised learning with multilingual pre-training has further accelerated progress in speech translation. Models such as SeamlessM4T and Whisper have been trained on vast amounts of multilingual speech data, demonstrating robust performance across multiple languages, even those with limited supervision. These architectures leverage shared latent representations across languages, facilitating cross-lingual transfer and improving translation accuracy. Despite these advancements, several challenges remain. Domain adaptation continues to be a key concern, as unsupervised models may struggle with domain-specific terminology and informal speech variations. Additionally, bias and fairness issues must be addressed to ensure that self-supervised models do not disproportionately favor high-resource languages at the expense of underrepresented dialects. Ethical considerations, including privacy concerns related to large-scale unsupervised data collection, must also be carefully managed to ensure the responsible deployment of these technologies. Looking ahead, unsupervised learning is expected to play a central role in the future of speech translation. As research continues to refine these methods, the potential for truly scalable and adaptable multilingual translation systems will increase. By reducing dependence on expensive human annotations, unsupervised speech translation models have the potential to democratize language technologies, making high-quality translation accessible for a broader range of languages and communities worldwide.

The Transformer-based model has been one of the most significant advancements in speech translation, revolutionizing the efficiency and accuracy of NMT and STT systems. The Transformer architecture, introduced by Vaswani et al.[26], replaced recurrent neural networks and long short-term memory networks as the dominant framework for language modeling and sequence-to-sequence tasks. With its self-attention mechanism, Transformers can model long-range dependencies more effectively, significantly improving translation fluency and coherence.

Adopting end-to-end Transformer models has accelerated performance gains in speech translation by eliminating the need for cascading ASR and text-based NMT pipelines. Systems such as Whisper[27] and SeamlessM4T[28] leverage Transformer architectures to directly map speech input to translated text or speech, minimizing error propagation and reducing latency. These models achieve state-of-the-art results in low-latency, high-quality translation tasks across multiple languages. Recent research has also introduced lightweight and efficient Transformer models, addressing the computational cost associated with standard Transformer architectures.

Techniques like quantization, knowledge distillation, and pruning have been applied to optimize these models for real-time speech translation applications, making them more suitable for mobile and edge devices deployment. Additionally, decoder-only architectures are emerging as an alternative to traditional encoder-decoder Transformers, further reducing inference time while maintaining competitive translation performance. Despite these advancements, several challenges remain. Transformer models require large-scale, high-quality datasets for training, which can be a limitation for low-resource languages. Bias and fairness in Transformer-based translations continue to be areas of concern, as models trained on

imbalanced datasets may exhibit skewed translation outputs. Moreover, the high computational cost of training large Transformer-based speech translation models raises questions about sustainability and accessibility, particularly for researchers and developers in resource-constrained environments. Looking forward, the integration of Transformer architectures with self-supervised learning and multimodal translation systems is expected to drive further improvements in speech translation. Multilingual Transformers that support zero-shot translation capabilities will continue to expand the scope of speech-to-speech and speech-to-text translation, making high-quality translation more accessible across diverse linguistic landscapes. As these models evolve, their application in real-time, low-latency, and domain-specific translation tasks will further enhance the usability and reliability of speech translation technologies.

**Table 2. Emerging and declined research themes**

| Year | End-to-End Systems | Neural Networks | Deep Learning | Zero-Shot Translation | Multimodal Translation | Semantic Learning | Real-Time Translation | Low-Resource Languages | Statistical Machine Translation (SMT) | Phrase-Based SMT | Rule-Based Translation | Offline Translation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2010 | 5 | 7 | 6 | 0 | 1 | 2 | 1 | 0 | 20 | 15 | 10 | 10 |
| 2011 | 7 | 8 | 8 | 0 | 2 | 3 | 2 | 1 | 18 | 13 | 8 | 8 |
| 2012 | 9 | 10 | 11 | 0 | 3 | 4 | 3 | 2 | 16 | 12 | 7 | 7 |
| 2013 | 12 | 12 | 14 | 1 | 4 | 5 | 4 | 3 | 14 | 11 | 6 | 6 |
| 2014 | 14 | 15 | 18 | 1 | 5 | 6 | 5 | 4 | 12 | 10 | 5 | 5 |
| 2015 | 17 | 18 | 22 | 2 | 6 | 8 | 6 | 5 | 10 | 8 | 4 | 4 |
| 2016 | 20 | 21 | 25 | 3 | 8 | 10 | 7 | 6 | 8 | 6 | 3 | 3 |
| 2017 | 23 | 25 | 30 | 4 | 10 | 12 | 9 | 7 | 6 | 5 | 2 | 2 |
| 2018 | 27 | 30 | 35 | 5 | 12 | 14 | 10 | 8 | 4 | 4 | 1 | 1 |
| 2019 | 31 | 35 | 40 | 6 | 14 | 16 | 11 | 9 | 3 | 3 | 1 | 0 |
| 2020 | 36 | 40 | 45 | 8 | 16 | 18 | 12 | 10 | 2 | 2 | 1 | 0 |
| 2021 | 40 | 45 | 50 | 10 | 18 | 20 | 14 | 12 | 1 | 1 | 0 | 0 |
| 2022 | 45 | 50 | 55 | 12 | 20 | 22 | 15 | 14 | 1 | 0 | 0 | 0 |
| 2023 | 50 | 55 | 60 | 14 | 22 | 24 | 16 | 15 | 0 | 0 | 0 | 0 |
| 2024 | 55 | 60 | 65 | 16 | 24 | 26 | 18 | 16 | 0 | 0 | 0 | 0 |

### *3.4. Collaborative Networks*

#### *3.4.1. Co-Authorship Analysis*

Several prominent researchers, such as Satoshi Nakamura, Alexander Waibel, and Shinji Watanabe, stand out due to their central positions and extensive networks. Nakamura's cluster is particularly noteworthy, involving collaborations with a diverse group of researchers from various institutions and countries. This suggests his leadership in coordinating large-scale, interdisciplinary projects, emphasizing the global nature of speech translation research. Waibel's extensive network similarly indicates his active participation in multiple research initiatives, reflecting his role in driving forward innovative solutions in speech translation.

The co-authorship map of speech translation research (Figure 8) highlights the intricate network of collaborations among researchers, offering insights into the field's collaborative dynamics and influential contributors. This visualization not only maps out the connections but also sheds light on the impact and productivity of these collaborative efforts. The map also reveals a high degree of diversity and specialization within the field. Researchers like Matteo Negri, Marcello Federico, and Jan Niehues are associated with distinct clusters, which suggests that they focus on specific subfields or methodologies within speech translation. This specialization is crucial for addressing different aspects of the technology, from foundational algorithm development to applied translation systems.

The presence of multiple specialized clusters highlights the field's multifaceted nature, with various groups pushing forward different aspects of speech translation research. Emerging researchers and new collaborations are also visible on the map, indicated by smaller node sizes and newer connections. This highlights the field's dynamic nature, with new contributors continuously integrating into existing networks.

These new collaborations often bring fresh perspectives and innovative approaches, essential for the field's growth and adaptation to new challenges. The extensive collaboration among researchers significantly enhances the quality and innovation of research in speech translation. Collaborative efforts allow for the pooling of diverse expertise and resources, leading to more comprehensive and robust research outcomes. The cross-pollination of ideas facilitated by these networks accelerates the development of new methodologies and technologies, pushing the boundaries of what is possible in speech translation.

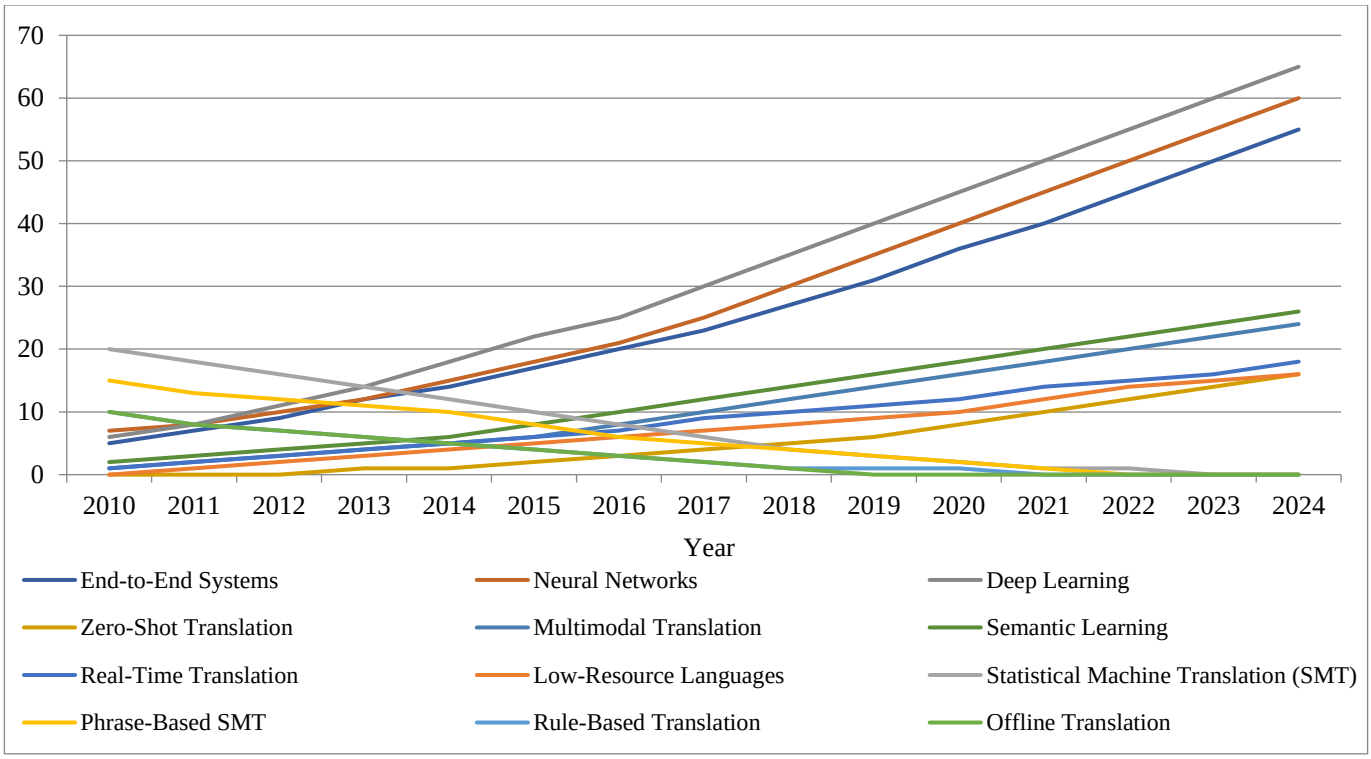


**Fig. 7 Emerging and declined research themes**

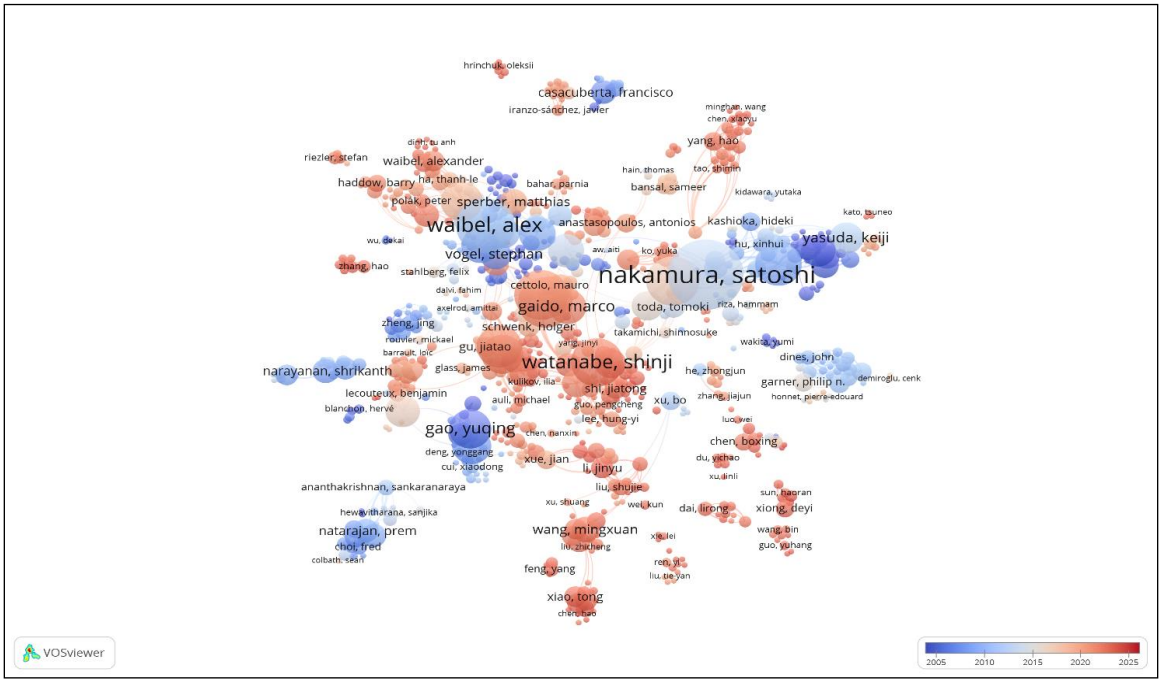


**Fig. 8 Co-authorship overlay map**

The presence of multiple specialized clusters highlights the field's multifaceted nature, with various groups pushing forward different aspects of speech translation research. Emerging researchers and new collaborations are also visible on the map, indicated by smaller node sizes and newer connections. This highlights the field's dynamic nature, with new contributors continuously integrating into existing networks. These new collaborations often bring fresh perspectives and innovative approaches, essential for the field's growth and adaptation to new challenges. The extensive collaboration among researchers significantly enhances the quality and innovation of research in speech translation. Collaborative efforts allow for the pooling

of diverse expertise and resources, leading to more comprehensive and robust research outcomes. The cross-pollination of ideas facilitated by these networks accelerates the development of new methodologies and technologies, pushing the boundaries of what is possible in speech translation. The influence of key contributors such as Nakamura, Waibel, and Watanabe is evident from their central positions in the co-authorship map. Their ability to attract and coordinate large collaborative efforts indicates their leadership and the respect they command within the research community. These influential researchers play a crucial role in setting research agendas, securing funding, and guiding the direction of the field. Their extensive networks are instrumental in tackling complex research problems that require interdisciplinary approaches.

*3.4.2. International Collaboration Patterns*

International collaborations significantly enhance the quality and innovation of research in speech translation. By pooling diverse expertise, resources, and perspectives, these collaborations lead to more comprehensive and innovative solutions. Countries like the United States and Japan, with their extensive networks, play a pivotal role in setting research agendas and facilitating knowledge exchange. This collaborative environment accelerates the development of new technologies and methodologies, driving the field forward.

The analysis of international collaboration patterns in speech translation, as depicted by the overlay map of co-authorship by countries (Figure 9), reveals a robust network of partnerships that drive the field forward. The map highlights key countries and their collaborative ties, providing insights into the global dynamics of research and development in this domain. The United States, Japan, China, and Germany emerged as central hubs in the international collaboration network. The size of the nodes representing these countries indicates their significant influence and prolific output in speech translation research.

The United States, in particular, stands out with numerous connections to other countries, reflecting its role as a major contributor and collaborator. Japan and Germany also show extensive networks, underlining their importance in fostering international research initiatives. The map reveals distinct regional clusters, with European countries like Germany, France, and the United Kingdom forming a tightly-knit network. These countries exhibit strong interconnections, suggesting frequent collaborations within Europe. Similarly, Asian countries such as Japan, China, and South Korea display a dense web of partnerships, highlighting regional cooperation. Cross-regional links between these clusters are also evident, with significant collaborations between the United States and both European and Asian countries. This cross-pollination of ideas and expertise is crucial for addressing the diverse challenges in speech translation. Emerging collaborations are indicated by newer and thinner lines connecting various countries. These lines suggest that countries like India, Brazil, and Malaysia increasingly participate in international research efforts. This trend points to the globalization of speech translation research, where new players are joining the established networks, contributing fresh perspectives and fostering innovation. The overlay map's color gradient, ranging from blue to red, indicates the evolution of these collaborations over time, with newer connections highlighted in red.

### *3.5. Journals and Conferences*

*3.5.1. Most Influential Journals and Conferences in the Field*

Referring to the table of the 10 most prolific journals in the field of speech translation (Table III), we can observe the significant influence these journals have on the dissemination and impact of research in this domain. Machine Translation and Computer Speech and Language are leading the list, each with 15 publications. These journals focus on both theoretical and practical advancements, addressing challenges and innovations in automated translation technologies. The relatively high citation counts for these journals, a total of 223 and 231, respectively, reflect their importance and the widespread influence of their published research. For instance, highly cited articles such as "Simultaneous translation of lectures and speeches" (72 citations) in Machine Translation and "MuST-C: A multilingual corpus for end-to-end speech translation" (100 citations) in Computer Speech and Language exemplify the impact of their contributions.

High-impact journals like IEEE Transactions on Audio, Speech and Language Processing and IEEE Signal Processing Magazine are particularly notable for their high citation concentrations. Despite having fewer publications (9 and 6, respectively), they have accumulated significant citations (295 and 341, respectively). This indicates that the research published in these journals often sets benchmarks in the field, providing foundational methods and comprehensive reviews that are widely referenced. For example, "The ATR multilingual speech-to-speech translation system" in IEEE Transactions on Audio, Speech and Language Processing has 108 citations, and "Making Machines Understand Us in reverberant rooms" in IEEE Signal Processing Magazine has 201 citations, highlighting their influential nature.

Interdisciplinary journals such as Lecture Notes in Computer Science and Applied Sciences (Switzerland) contribute to the broader research context by integrating speech translation with other scientific domains. Although these journals have moderate citation counts, their role in disseminating research that spans multiple disciplines is crucial. For example, "Phrase-based statistical machine translation" in Lecture Notes in Computer Science has 212 citations, reflecting its broad impact across computational linguistics and machine learning. Similarly, Applied Sciences Switzerland), with its practical focus, contributes valuable insights through articles like "Cascade or Direct Speech Translation? A Case Study in Chinese-English Translation" with 5 citations. A prominent

trend observed in the citation patterns is the high citation rates for articles addressing end-to-end translation systems and multilingual corpora. These topics are critical as they address the scalability and effectiveness of speech translation technologies in real-world applications. The highly cited article "MuST-C: A multilingual corpus for end-to-end speech translation", with 100 citations, is a testament to the research community's focus on creating comprehensive datasets and robust translation systems. Evaluative studies and benchmarking efforts also attract significant citations, highlighting the importance of empirical performance assessments and system comparisons in advancing the field. Journals that publish findings from major evaluation campaigns, such as the Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) and ICASSP, play a crucial role in this regard. For instance, "Findings of the IWSLT 2020 evaluation campaign", with 97 citations, underscores the value of these evaluative studies.

Practical applications and real-world implementations are emphasized in journals like IEICE Transactions on Information and Systems and the Journal of the National Institute of Information and Communications Technology. Although these journals show more focused citation patterns, they highlight the practical impact and innovation within applied settings.

Articles such as "Development of the 'VoiceTra' multilingual speech translation system" with 3 citations and "Multilingual speech synthesis system" with 8 citations demonstrate the importance of applied research in driving technological advancements and adoption. The analysis of the top 10 most prolific conferences in the field of speech translation (Table IV) reveals significant insights into the dissemination and impact of research within this domain. Leading the list is the Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, with an impressive 150 publications and 1687 total citations. This conference's high publication count and citation volume underscore its role as a premier venue for presenting cutting-edge research and significant advancements in speech translation. Similarly, the ICASSP, IEEE International Conference on Acoustics, Speech, and Signal Processing - Proceedings follows closely with 119 publications and 2099 total citations, indicating its prominence as a key forum for disseminating impactful research.

Articles such as "Sequence-to-sequence models can directly translate foreign speech" (203 citations) and "Recent developments on ESPNeT toolkit boosted by Conformer" (164 citations) highlight the conferences' focus on innovative models and toolkits that enhance speech processing capabilities.

The Proceedings of the Annual Meeting of the Association for Computational Linguistics, ACL, and various International Workshop on Spoken Language Translation (IWSLT) conferences emphasize the importance of evaluation and benchmarking in advancing the field. With 69 publications and 990 total citations for ACL and notable contributions from IWSLT conferences such as "Findings of the IWSLT 2020 evaluation campaign" (97 citations), these venues underscore the critical role of empirical performance assessments and system comparisons.

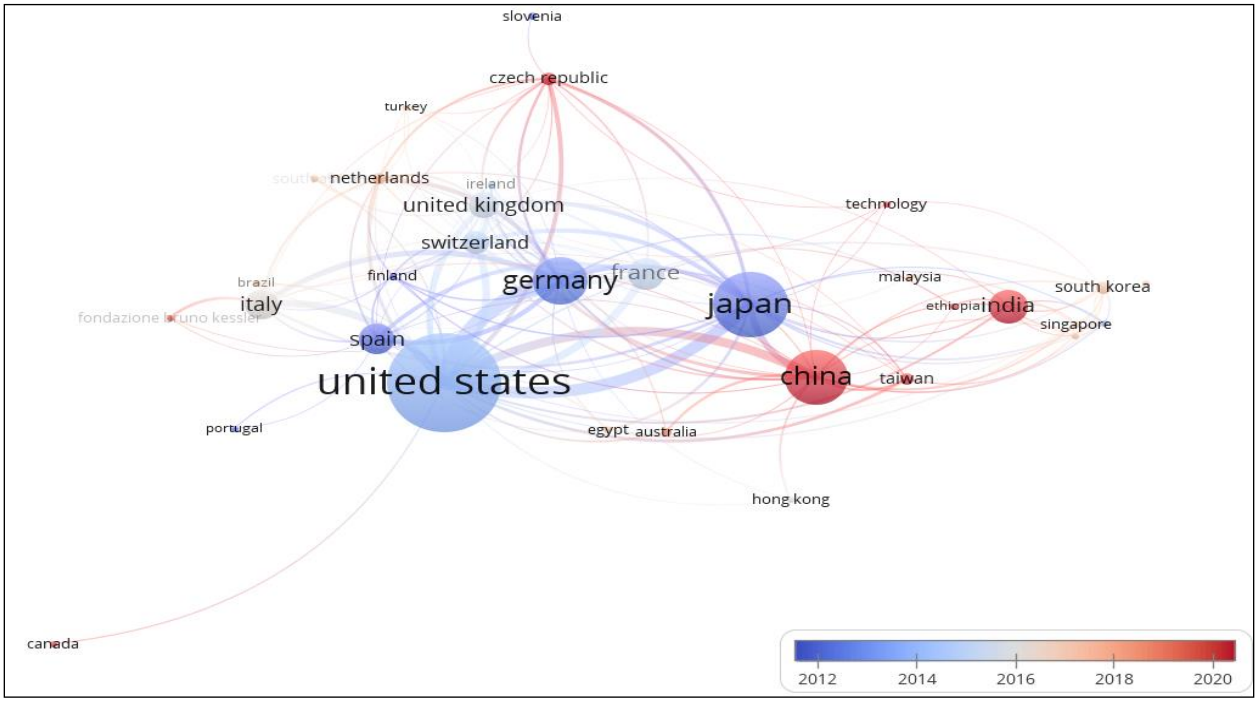


**Fig. 9 Co-authorship by countries overlay map**

**Table 3. 10 Most prolific journals in the field of speech translation**

| # | Journal | Publication Count | Total Citations | Highly Cited Article Title | Citations |
|---|---|---|---|---|---|
| #1 | Machine Translation | 15 | 223 | Simultaneous translation of lectures and speeches [29] | 72 |
| #2 | Computer Speech and Language | 15 | 231 | MuST-C: A multilingual corpus for end-to-end speech translation[18] | 100 |
| #3 | IEEE Transactions on Audio, Speech and Language Processing | 9 | 295 | The ATR multilingual speech-to-speech translation system[30] | 108 |
| #4 | Speech Communication | 8 | 107 | Talker discrimination across languages[31] | 35 |
| #5 | IEEE/ACM Transactions on Audio Speech and Language Processing | 8 | 76 | End-to-End Speech Translation with Transcoding by Multi-Task Learning for Distant Language Pairs[32] | 27 |
| #6 | IEICE Transactions on Information and Systems | 8 | 13 | Development of the "VoiceTra" multilingual speech translation system[33] | 3 |
| #7 | Journal of the National Institute of Information and Communications Technology | 6 | 10 | Multilingual speech synthesis system[34]vv | 8 |
| #8 | IEEE Signal Processing Magazine | 6 | 341 | Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition[22] | 201 |
| #9 | Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) | 5 | 6 | Towards the development of the multilingual multimodal virtual agent[35] | 4 |
| #10 | Applied Sciences (Switzerland) | 4 | 14 | Cascade or Direct Speech Translation? A Case Study[4] | 5 |

Lecture Notes in Computer Science and Applied Sciences (Switzerland) further contribute to the field by integrating speech translation with broader research areas, reflecting their interdisciplinary impact.

Despite moderate citation counts, their contributions to foundational research and practical applications are significant, as seen in articles like "Phrase-based statistical machine translation" (212 citations) and "Cascade or Direct Speech Translation? A Case Study in Chinese-English Translation" (5 citations). A prominent trend observed in citation patterns is the high citation rates for articles addressing end-to-end translation systems and multilingual corpora, reflecting the research community's focus on creating comprehensive datasets and robust translation systems. Additionally, conferences like the 2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2023), focusing on architectural advancements such as decoder-only architectures, highlight the importance of innovative system designs. Articles like "On Decoder-Only Architecture for Speech-to-Text Translation" (8 citations) emphasise improving the efficiency and accuracy of translation systems.

*3.5.2 Analysis of Publication Outlets*

Leading journals such as Machine Translation and Computer Speech and Language are pivotal in publishing both theoretical and practical advancements in automated translation technologies. These journals emphasize research that addresses core challenges and introduces innovative solutions in speech translation.

High citation counts in these journals indicate their influential role in the academic community. For instance, the article "MuST-C: A multilingual corpus for end-to-end speech translation" in Computer Speech and Language demonstrates the importance of creating robust multilingual datasets for advancing translation systems.

IEEE Transactions on Audio, Speech and Language Processing and IEEE Signal Processing Magazine are recognized for their rigorous peer review and high-impact articles that often set new standards in the field. The high citation volumes for these journals reflect their role in disseminating foundational research and comprehensive reviews widely referenced by other scholars. Articles like "The

ATR multilingual speech-to-speech translation system" illustrate the significant contributions of these journals to the advancement of speech translation technologies. Conferences like INTERSPEECH and ICASSP are leading forums for presenting cutting-edge research and technological advancements. These conferences not only attract a high number of submissions but also feature highly cited papers that drive the research agenda. For example, "Sequence-to-Sequence models can directly translate foreign speech," presented at INTERSPEECH, highlights the conference's focus on innovative models that enhance speech translation capabilities. The Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) and various International Workshop on Spoken Language Translation (IWSLT) conferences are crucial for presenting evaluative studies and benchmarking efforts. These conferences emphasize the importance of empirical performance assessments and comparisons, as seen in the highly cited "Findings of the IWSLT 2020 evaluation campaign". Such studies are vital for advancing the field through rigorous evaluation and validation of new approaches. The high citation rates for end-to-end translation systems and multilingual corpora indicate the research community's focus on scalability and effectiveness in real-world applications. Evaluative studies and benchmarking, particularly in conferences like ACL and IWSLT, underscore the field's emphasis on performance assessment and system improvements. Technological advancements and toolkits presented at conferences like ICASSP highlight the ongoing innovation in speech processing capabilities. Interdisciplinary contributions from journals such as Lecture Notes in Computer Science and practical applications from journals like Applied Sciences (Switzerland) demonstrate the integration of speech translation with broader research areas and real-world implementations. This interdisciplinary approach helps bridge gaps between different fields and fosters innovative solutions.

**Table 4. 10 Most prolific conferences in the fields of speech translation**

| # | Conference | Publication Count | Total Citations | Highly Cited Article Title | Citations |
|---|---|---|---|---|---|
| #1 | Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH | 150 | 1687 | Sequence-to-sequence models can directly translate foreign speech[21] | 203 |
| #2 | ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings | 119 | 2099 | Recent developments on ESPNeT toolkit boosted by conformer[24] | 164 |
| #3 | Proceedings of the Annual Meeting of the Association for Computational Linguistics | 69 | 990 | Findings of the IWSLT 2020 evaluation campaign[36] | 97 |
| #4 | Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) | 49 | 354 | Phrase-based statistical machine translation[20] | 212 |
| #5 | 20th International Conference on Spoken Language Translation, IWSLT 2023 - Proceedings of the Conference | 37 | 91 | Findings of the IWSLT 2023 Evaluation Campaign[37] | 40 |
| #6 | IWSLT 2021 - 18th International Conference on Spoken Language Translation, Proceedings | 25 | 187 | Findings of the IWSLT 2021 Evaluation Campaign[38] | 62 |
| #7 | IWSLT 2022 - 19th International Conference on Spoken Language Translation, Proceedings of the Conference | 21 | 241 | Findings of the IWSLT 2022 Evaluation Campaign[39] | 82 |
| #8 | 2023 IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2023 | 13 | 21 | On Decoder-Only Architecture For Speech-to-Text and Large Language Model Integration[40] | 8 |
| #9 | EUROSPEECH 2003 - 8th European Conference on Speech Communication and Technology | 12 | 225 | Creating corpora for speech-to-speech translation[41] | 121 |
| #10 | 8th International Conference on Spoken Language Processing, ICSLP 2004 | 11 | 79 | Using word lattice information for a tighter coupling in speech translation systems[42] | 38 |

*3.5.3 Societal Implications and Ethical Considerations in Speech Translation Technologies*

The rapid advancements in speech translation technologies have significantly impacted various domains, including international communication, accessibility, and multilingual information dissemination. By enabling real-time translation between languages, these systems facilitate seamless interaction in business, diplomacy, education, and healthcare, among other fields. The transition from traditional statistical models to end-to-end neural architectures has further improved translation fluency and efficiency, making these technologies more practical for real-world applications. However, the integration and widespread deployment of speech translation models introduce critical societal implications and ethical challenges that must be carefully considered to ensure their responsible and equitable use.

One of the most evident benefits of speech translation technologies is their role in reducing linguistic barriers. By providing instantaneous translations, they enhance cross-cultural communication and enable individuals to engage in global interactions regardless of their linguistic background. This has substantial implications for international collaboration, particularly in regions with high linguistic diversity. Additionally, these technologies improve accessibility for individuals with hearing impairments and non-native speakers, allowing them to participate in multilingual conversations. Another crucial aspect is their potential in language preservation, particularly for low-resource and endangered languages. Through speech-to-text and speech-to-speech translation, these systems can contribute to the digital documentation and revitalization of lesser-spoken languages, ensuring their continued presence in technological and academic domains. However, there is a risk that current models primarily focus on high-resource languages, which may contribute to the further marginalization of underrepresented languages if not explicitly designed to support linguistic diversity.

Despite these advantages, speech translation technologies raise significant concerns related to fairness and bias. Neural models are trained on large-scale datasets that may reflect existing linguistic and cultural biases, potentially leading to translation inaccuracies or discriminatory outputs. Variations in pronunciation, dialects, and regional accents can result in performance disparities, disproportionately affecting speakers from specific linguistic backgrounds. If these biases are not adequately addressed, they could reinforce existing social inequalities and limit the usability of speech translation systems for diverse populations. Ensuring fairness in translation outputs requires the development of more inclusive datasets and continuous evaluation of model performance across different demographic groups. Privacy and data security are also fundamental considerations in deploying speech translation technologies. Many contemporary models rely on cloud-based processing, which involves transmitting and storing speech data on remote servers. This raises concerns about data confidentiality, unauthorized access, and potential misuse of personal information. Data breaches or surveillance risk becomes particularly critical in contexts where sensitive conversations are translated, such as medical consultations or legal proceedings. Ethical deployment necessitates implementing stringent data protection measures, including encryption protocols, anonymization techniques, and transparent consent mechanisms that allow users to control the storage and usage of their speech data. Adherence to international data privacy regulations, such as the General Data Protection Regulation (GDPR), is essential to ensuring user trust and ethical compliance.

Another key consideration is the impact of speech translation automation on professional interpreters and translators. While these technologies offer efficiency and scalability, they may reduce the demand for human linguistic experts in various industries. Professional interpreters play a crucial role in contexts requiring linguistic precision, cultural sensitivity, and contextual adaptation—such as legal settings, diplomatic negotiations, and high-stakes medical translations. The increasing reliance on automated systems should, therefore, be approached with a balanced perspective, where human expertise remains integral to refining and overseeing machine-generated translations. Hybrid models that combine AI-driven translation with human post-editing and quality control can help maintain translated content's reliability and contextual appropriateness.

Furthermore, the capability of speech translation technologies to generate real-time translated speech raises ethical concerns regarding misinformation and digital manipulation. Advances in speech synthesis and translation models have the potential to be exploited for deepfake audio generation, where fabricated translations or manipulated speech recordings could be used to mislead audiences, spread false information, or distort public discourse. The growing sophistication of these models necessitates the development of robust verification mechanisms, watermarking techniques, and regulatory policies to prevent misuse while ensuring the authenticity of translated content. Given these considerations, the development and deployment of speech translation technologies must be guided by ethical principles that prioritize fairness, transparency, and accountability. Addressing biases, ensuring data privacy, promoting linguistic inclusivity, and mitigating the impact on professional translators are essential for fostering the responsible advancement of these technologies. As speech translation continues to evolve, interdisciplinary collaboration among linguists, AI researchers, policymakers, and ethicists will be crucial in shaping frameworks that maximize societal benefits while minimizing ethical risks. Through thoughtful design and regulatory oversight, speech translation technologies can serve as powerful tools for enhancing global communication while upholding ethical and equitable standards.

## 4. Conclusion

This comprehensive bibliometric analysis of speech translation research from 2000 to 2024 provides valuable insights into the field's evolution, trends, and key contributions. By leveraging data from the Scopus database, we identified the most prolific authors, institutions, countries, and publication outlets, as well as the prevailing research themes and collaboration patterns. Our findings indicate significant growth in speech translation research, particularly over the last decade, driven by advancements in machine learning, deep learning, and neural networks. Leading journals such as Machine Translation, Computer Speech and Language, and IEEE Transactions on Audio, Speech, and Language Processing have been instrumental in publishing groundbreaking research that addresses both theoretical foundations and practical applications. High citation counts in these journals highlight their influence and the importance of their contributions to the field.

Conferences like INTERSPEECH and ICASSP have emerged as premier venues for presenting cutting-edge research and significant technological advancements in speech translation. These conferences not only attract high numbers of publications but also feature highly cited papers that set new standards and drive the research agenda. The focus on empirical performance assessments and benchmarking at conferences such as ACL and IWSLT underscores the critical role of rigorous evaluation in advancing speech translation technologies. Key trends observed in citation patterns reveal a growing emphasis on end-to-end translation systems and multilingual corpora, reflecting the research community's efforts to develop scalable and effective technologies for real-world applications. The shift from traditional statistical and rule-based methods to advanced neural and machine learning techniques indicates a maturing field that continues to innovate and evolve. The collaborative nature of speech translation research is evident from extensive co-authorship networks and international partnerships. Leading researchers and institutions from technologically advanced countries have made substantial contributions, driving innovation and addressing complex challenges through collaborative efforts. The integration of speech translation with broader research areas, as highlighted by interdisciplinary journals and conferences, further enhances the field's development and application. Overall, this bibliometric analysis highlights speech translation research's dynamic and multi-faceted nature. The ongoing contributions from high-impact journals and conferences, coupled with collaborative and interdisciplinary efforts, will continue to play a crucial role in shaping the future of speech translation technologies. These advancements are essential for promoting global communication, enhancing accessibility, and bridging language barriers in an increasingly interconnected world.

## References


[1] Peter F. Brown et al., "A Statistical Approach To Machine Translation," *Computational Linguistics*, vol. 16, no. 2, pp. 79-85, 1990. [Google Scholar] [Publisher Link]

[2] Yonghui Wu et al., "Google's Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation," *arXiv*, pp. 1-23, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[3] Nivedita Sethiya, and Chandresh Kumar Maurya, "End-to-End Speech-to-Text Translation: A Survey," *arXiv*, pp. 1-75, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Thierry Etchegoyhen et al., "Cascade or Direct Speech Translation? A Case Study," *Applied Sciences*, vol. 12, no. 3, pp. 1-24, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] L. Bentivogli et al., "Cascade Versus Direct Speech Translation: Do the Differences Still Make A Difference?," *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, vol. 1, pp. 2873-2887, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Ye Jia et al., "Direct Speech-to-Speech Translation with A Sequence-to-Sequence Model," *arXiv*, pp. 1-5, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[7] Jan Niehues et al., "Tutorial: End-to-End Speech Translation," *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts*, pp. 10-13, 2021. [CrossRef] [Publisher Link]

[8] Parnia Bahar, Tobias Bieschke, and Hermann Ney, "A Comparative Study on End-to-End Speech to Text Translation," *IEEE Automatic Speech Recognition and Understanding Workshop*, Singapore, pp. 792-799, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[9] Sameer Bansal et al., "Low-Resource Speech-to-Text Translation," *arXiv*, pp. 1-5, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[10] Sameer Bansal et al., "Towards Speech-to-Text Translation without Speech Recognition," *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, vol. 2, pp. 474-479, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[11] Hirofumi Inaguma et al., "Multilingual End-to-End Speech Translation," *IEEE Automatic Speech Recognition and Understanding Workshop*, Singapore, pp. 570-577, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[12] Yuchen Liu et al., "Synchronous Speech Recognition and Speech-to-Text Translation with Interactive Decoding," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 5, pp. 8417-8424, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Yichao Du et al., "Regularizing End-to-End Speech Translation with Triangular Decomposition Agreement," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 10, pp. 10590-10598, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Gerard I. Gállego et al., "End-to-End Speech Translation with Pre-trained Models and Adapters: {UPC} at {IWSLT} 2021," *Proceedings of the 18th International Conference on Spoken Language Translation (IWSLT)*, pp. 110-119, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] Xuancai Li et al., "End-to-End Speech Translation with Adversarial Training," *Proceedings of the First Workshop on Automatic Simultaneous Translation*, pp. 10-14, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Zhihao Zhou et al., "Sign-to-Speech Translation Using Machine-Learning-Assisted Stretchable Sensor Arrays," *Nature Electronics*, vol. 3, no. 9, pp. 571-578, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] Shigeki Karita et al., "A Comparative Study on Transformer vs RNN in Speech Applications," *IEEE Automatic Speech Recognition and Understanding Workshop*, Singapore, pp. 449-456, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[18] Roldano Cattoni et al., "MuST-C: A Multilingual Corpus for End-to-End Speech Translation," *Computer Speech & Language*, vol. 66, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Philipp Koehn et al., "Edinburgh System Description for the 2005 IWSLT Speech Translation Evaluation," *Proceedings of the Second International Workshop on Spoken Language Translation*, 2005. [Google Scholar] [Publisher Link]

[20] Richard Zens, Franz Josef Och, and Hermann Ney, "Phrase-Based Statistical Machine Translation," *KI: Advances in Artificial Intelligence*, pp. 18-32, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[21] Ron J. Weiss et al., "Sequence-to-Sequence Models Can Directly Translate Foreign Speech," *arXiv*, pp. 1-5, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[22] Takuya Yoshioka et al., "Making Machines Understand Us in Reverberant Rooms: Robustness Against Reverberation for Automatic Speech Recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 114-126, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[23] Toshiyuki Takezawa et al., "Toward A Broad-Coverage Bilingual Corpus for Speech Translation of Travel Conversations in the Real World," *Proceedings of the Third International Conference on Language Resources and Evaluation*, pp. 147-152, 2002. [Google Scholar] [Publisher Link]

[24] Pengcheng Guo et al., "Recent Developments on Espnet Toolkit Boosted By Conformer," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, ON, Canada, pp. 5874-5878, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] B. Bangalore, G. Bordel, and G. Riccardi, "Computing Consensus Translation From Multiple Machine Translation Systems," *IEEE Workshop on Automatic Speech Recognition and Understanding*, Madonna di Campiglio, Italy, pp. 351-354, 2001. [CrossRef] [Google Scholar] [Publisher Link]

[26] Ashish Vaswani et al., "Attention is all You Need," *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Google Scholar] [Publisher Link]

[27] Alec Radford et al., "Robust Speech Recognition Via Large-Scale Weak Supervision," *Proceedings of the 40th International Conference on Machine Learning*, Honolulu, Hawaii, USA, pp. 28492-28518, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] Loïc Barrault et al., "SeamlessM4T-Massively Multilingual & Multimodal Machine Translation," *arXiv*, pp. 1-111, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[29] Christian Fügen, Alex Waibel, and Muntsin Kolss, "Simultaneous Translation of Lectures and Speeches," *Machine Translation*, vol. 21, no. 4, pp. 209-252, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[30] S. Nakamura et al., "The ATR Multilingual Speech-to-Speech Translation System," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 365-376, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[31] Mirjam Wester, "Talker Discrimination Across Languages," *Speech Communication*, vol. 54, no. 6, pp. 781-790, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[32] Takatomo Kano, Sakriani Sakti, and Satoshi Nakamura, "End-to-End Speech Translation with Transcoding by Multi-Task Learning for Distant Language Pairs," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1342-1355, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[33] Shigeki Matsuda et al., "Development of the 'VoiceTra' Multi-Lingual Speech Translation System," *IEICE Transactions on Information System*, vol. E100-D, no. 4, pp. 621-632, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[34] Yoshinori Shiga, and Hisashi Kawai, "Multilingual Speech Synthesis System," *Journal of the National Institute of Information and Communications Technology*, vol. 59, no. 3.4, pp. 21-28, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[35] Inese Vīra, Jānis Teseļskis, and Inguna Skadiņa, "Towards the Development of the Multilingual Multimodal Virtual Agent," *Advances in Natural Language Processing*, pp. 470-477, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[36] Ebrahim Ansari et al., "Findings of the IWSLT 2020 Evaluation Campaign," *Proceedings of the 17th International Conference on Spoken Language Translation*, pp. 1-34, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[37] Milind Agarwal et al., "Findings of the Iwslt 2023 Evaluation Campaign," *Proceedings of the 20th International Conference on Spoken Language Translation*, pp. 1-61, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[38] Antonios Anastasopoulos et al., "Findings of the Iwslt 2021 Evaluation Campaign," *Proceedings of the 18th International Conference on Spoken Language Translation*, pp. 1-29, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[39] Antonios Anastasopoulos et al., "Findings of the IWSLT 2022 Evaluation Campaign," *Proceedings of the 19th International Conference on Spoken Language Translation*, pp. 98-157, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[40] Jian Wu et al., "On Decoder-Only Architecture For Speech-to-Text and Large Language Model Integration," *IEEE Automatic Speech Recognition and Understanding Workshop*, Taipei, Taiwan, pp. 1-8, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[41] Genichiro Kikui et al., "Creating Corpora for Speech-to-Speech Translation," *8th European Conference on Speech Communication and Technology*, pp. 381-384, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[42] Tanja Schultz et al., "Using Word Latice Information for A Tighter Coupling in Speech Translation Systems," *Interspeech, 8th International Conference on Spoken Language Processing ICC Jeju*, Jeju Island, Korea, pp. 41-44, 2004. [CrossRef] [Google Scholar] [Publisher Link]