

# **Voices of Europe: Mapping Cultural and Linguistic Diversity in Eurovision Song Lyrics**

Report on a dataset of Eurovision songs from a linguistic perspective

**TechRhythm:** Elīza Anna Actiņa, Jeļizaveta Larionova, Amanda Sprukte

Riga Technical University,  
Faculty of Computer Science, Information Technology and Energy,  
Digital Humanities master's programme,  
Introduction to Digital Humanities course (DE0802)

Riga, 2025

## **INTRODUCTION**

Since its establishment in 1956, the Eurovision Song Contest (ESC) has functioned as more than a musical competition. It is a recurring transnational event in which participating countries present linguistic, cultural and symbolic identities to an international audience. Language choice in song lyrics plays a particularly visible role in this process, as it reflects tensions between national identity, cultural proximity and international accessibility. As such, Eurovision provides a valuable corpus for examining how language operates within popular culture and how linguistic choices relate to broader geopolitical and cultural dynamics.

While existing research has often focused on voting behavior or political alliances, fewer studies have conducted systematic, large-scale analyses of song lyrics and thematic content across multiple languages and over time. From a digital humanities perspective, Eurovision song lyrics offer an opportunity to combine computational analysis with cultural interpretation. In this project, language families are used as a theoretical framework, as they represent shared linguistic heritage and historical connections between nations.

The overarching research question guiding this project is: how do the Eurovision Song Contest participants' song lyrics and linguistic choices reflect geopolitical and cultural ties between the nations?

To address this question, our project investigates four interconnected sub-questions:

1. What similarities and differences exist among the topics and vocabularies of song lyrics in languages of shared language families?
2. How have the language choice and topic/vocabulary trends of each contest's top five most highly rated songs changed across the 2008-2025 timeframe?
3. How are different languages used together in bilingual or multilingual songs, particularly with English functioning as a lingua franca on the international stage?
4. How have participating countries used other nations' languages in their songs?

Accordingly, hypotheses were outlined for each sub-question:

1. Songs within the same language family will share greater lexical and thematic similarity than songs across different families.
2. Highly ranked songs will display more standardized language choices and thematic patterns over time.

3. English will most often appear alongside national languages in multilingual songs as a means of increasing accessibility rather than replacing local identity.
4. The use of non-native languages will reflect cultural affiliation or regional proximity between countries.

The project examines 724 songs across an 18 year period from 2008 to 2025, corresponding to the introduction of the two-semifinal and one-final contest format, which stabilized the participation and competition structure. Although songs were selected for the 2020 contest, the event was cancelled due to the COVID-19 pandemic. Because this year lacks performance, voting and ranking data, it has been excluded from the top five songs' text analysis section, which relies on this data.

Methodologically, the study combines curated lyric datasets with digital text analysis using Voyant Tools, Tableau Public, Gephi and QGIS. Lyrics are grouped by language families and branches and analyzed through word frequency, contextual usage and geospatial trends. A distinction is made between surface-level lyrical vocabulary and topic-based semantic vocabularies, allowing for a structured comparison of meaning. The main structure of our combined digital dataset was built in Excel by combining relevant data from two existing datasets: *Lyrics of every Eurovision song in the history* (curator: minitree), and *The Eurovision Song Contest Dataset* (curator: Janne Spijkervet). The former primarily provides data on the general song attributes, such as source country, year, languages, lyrics and their English translations, while the latter offers structured public response data of rankings and scores for each relevant song. We filtered both datasets to include only songs within our timeframe from 2008 to 2025, resulting in a combined dataset of 724 songs. This was then supplemented by linking language metadata with every language's family and family branch. The data also had to be checked and cleaned for data validity before analysis, e.g. to remove non-lyrical content such as [Chorus] and [Verse] from the lyrics' data fields and to merge fields with same values, but different names, such as Czechia and Czech Republic under participating countries. For analysis, we used digital tools like Voyant Tools, Tableau Public, Gephi and QGIS; for the geospatial map visualization, the country polygon data was sourced from the [Natural Earth](#) public domain map dataset and displayed on this website with the help of the Leaflet JavaScript library. The website itself was built using GitHub and GitHub Pages.

This project does not claim to identify the lyrical or linguistic intentions of individual artists or national broadcasters. Lyrics vary in length, structure and translation quality, and multilingual songs complicate categorization. Frequency-based methods privilege repetition over nuance, particularly in smaller language groups. The findings should thus be understood as patterns and tendencies rather than definitive explanations of cultural or political motivation.

Overall, this project explores how digital humanities' methods can illuminate linguistic and cultural dynamics within Eurovision, offering a structured and comparative perspective on how language functions in one of Europe's most visible cultural events.

## **TEXT ANALYSIS ACROSS LANGUAGE FAMILIES**

Our first research sub-question investigates: What similarities and differences exist among the topics and vocabularies of song lyrics in languages of shared language families? Analyzing a corpus of Eurovision song lyrics from 2008 to 2025, this analysis compares lexical repetition (lyrics) and semantic focus (topics) across four groups: all language families combined, Indo-European languages, Uralic languages and other language families (such as Afro-Asiatic, Kartvelian and Turkic). We use Voyant Tools (Cirrus, Terms, Contexts and Trends) to move from surface-level word frequency to broader thematic patterns.

### **Shared vocabulary across all language families**

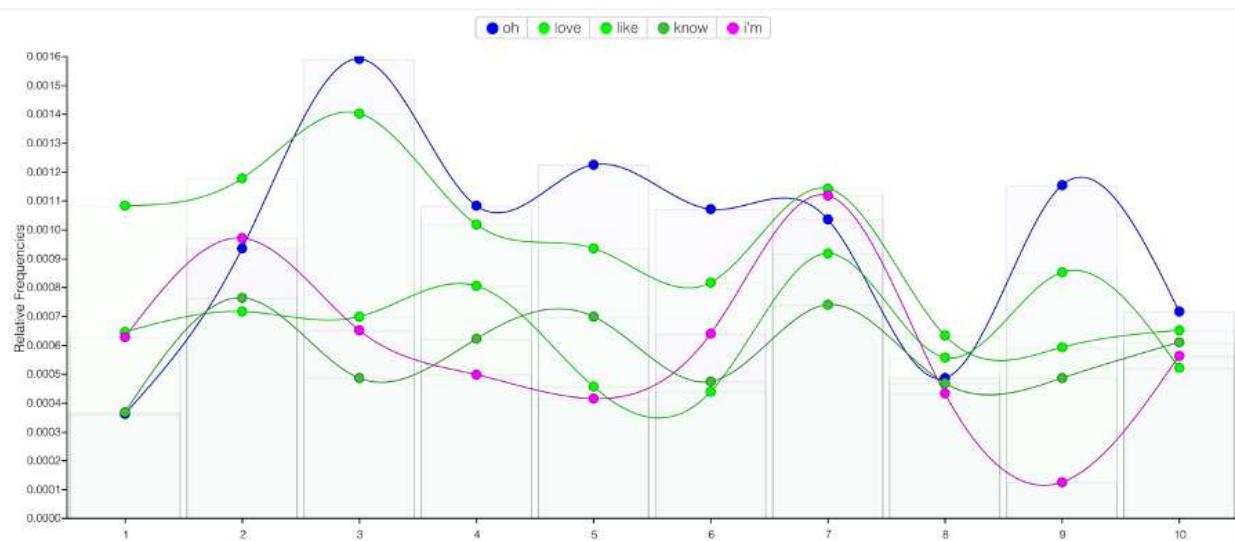
As seen in this Cirrus visualization, a small set of high-frequency words dominates Eurovision lyrics across all languages, indicating a shared performative vocabulary.



	Term	Count
1	oh	1632
2	love	1597
3	you	1116
4	I'm	1020
5	know	965
6	I'm	746
7	just	733
8	let	636
9	it's	591
10	come	591
11	time	571
12	heart	568
13	life	538
14	want	532
15	got	501
16	feel	501
17	yeah	499
18	make	492
19	gonna	474

According to the Terms panel, the top five words - such as *oh* (1632), *love* (1597), *like* (1116), *I'm* (1020) and *know* (965) - appear 6330 times, together accounting for a substantial share of all word occurrences. Contexts' analysis shows that many of these words function as fillers or rhythmic markers rather than carriers of meaning, appearing repeatedly in similar lyrical positions.

Trends' visualization confirms that these words remain prominent across most of the 18 years analyzed, indicating long-term stability rather than short-lived trends.



This suggests that Eurovision lyrics across language families rely on a shared performative vocabulary shaped by pop music conventions rather than linguistic structure alone.

### Shared topics across all language families: meaning beyond sound

While lyrics tend to rely on word repetition, when the analysis shifts from lyrics to topics, the vocabulary becomes more semantically focused across all Eurovision songs.



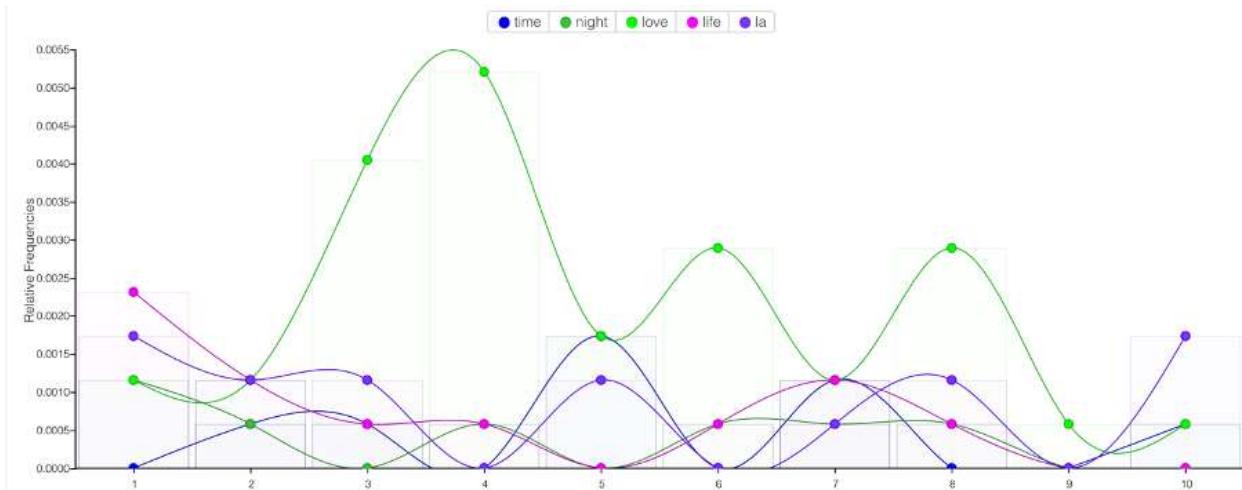
In the topic-based Cirrus visualization, words such as *love*, *life*, *time* and *night* emerge as the most prominent. The Terms panel shows that *love* alone appears 37 times, making it the most common thematic concept across the entire dataset.

	Term	Count
1	love	37
2	la	15
3	life	12
4	time	8
5	night	7
6	je	7
7	heart	7
8	dance	6
9	song	5
10	shine	5
11	les	5
12	lights	5
13	day	5
14	angel	5
15	storm	4
16	running	4
17	run	4
18	oh	4
19	kiss	4

Unlike lyrical fillers, Contexts' analysis demonstrates that topic words appear in varied but meaningful constructions, often forming the emotional or narrative core of songs.

Left	Term ↑	Right
סאלט אן אס (Salt N Pepa) A Century of	love	Complice Day After Day Vrag
Vodka Femele fatale Let Me	love	You Senhora do mar (Negras
Firefly Let' tmou Believe Again	love	Symphony (Simfonija) Dance with Me
Always This Is Our Night	love	Hora din Moldova Carry Me
td>Čaroban (Чаробан) Get You In	love	for a While One More
Running Scared Watch My Dance	love	in Rewind The Secret Is
in Rewind The Secret Is	love	Never Alone With Love Baby
Is Love Never Alone With	love	Baby I'm Still Alive Angel
One Change Rockefeller Street I	love	Belarus Angel in Disguise New
Tomorrow Lipstick Sognu Madness of	love	(Follia d'amore) I Can Taken
Uh – Oh Oh) La La	love	Should've Known Better Sound of
Vida minha Be My Guest	love	Unlimited Verjamem Nebo Euphoria I'm
Nebo Euphoria I'm a Joker	love	Me Back Don't Close Your
Eyes Stay Korake ti znam	love	Is Blind Love Will Set
ti znam Love Is Blind	love	Will Set You Free Echo
L'amore è femmina (Out of	love	) When the Music Dies Quédate
Standing Still Shine Straight into	love	Mižerja Only Teardrops What If

Trends show that these thematic concepts remain stable over time, suggesting that Eurovision maintains a consistent set of shared topics even as languages vary.



## **Indo-European languages: strong standardization in lyrics and topics**

Isolating only songs using languages from Indo-European language families, the emerging pattern is one of high vocabulary repetition and stable thematic focus across time.



Term	Count
oh	1632
um	1587
uh	1118
um	1069
knew	966
um	746
just	733
uh	636
is	561
comes	551
tens	571
heart	568
like	556
want	552
got	551
feel	501
yeah	496
nailed	482
gonna	474

	Term	Count
20	it's	4
21	goodbye	4
22	its	4
23	you	4
24	any	4
25	alive	4
26	way	3
27	walk	3
28	stay	3
29	stars	3
30	stand	3
31	rise	3
32	new	3
33	man	3
34	you	3
35	laptop	3
36	oh	3
37	I'm	3
38	home	3

	Term	Count
39	hold	3
40	hasta	3
41	gravity	3
42	friend	3
43	you	3
44	feel	3
45	eyes	3
46	et	3
47	et	3
48	et	3
49	believe	3
50	oh	2

The Indo-European language group displays the highest degree of standardization. In lyrical analysis, Cirrus and Terms tools reveal that a very small number of words dominate overall vocabulary. The top exclamation word *oh* appears 1632 times, while the top five words together account for 6330 occurrences within the top 50 terms; Contexts' analysis confirms that these words are used in repetitive, predictable positions.

Left	Term	Right
Save us from all tears	oh	Lord, save us... You can
Och, don't wanna be lonely	oh	no no? Why won't anybody
written in your eyes And	oh	, we made it! We finally
call For Life For Life	oh	, I come from a nation
how to write a song	oh	Europe, where oh where did
a song On Europe, where	oh	where did it all go
Aid don't forget the Swiss (	oh	Casanova, -nova, -nova) A thousand
When I think of you	oh	Casanova I feel love would
feel it in your heart?	oh	Casanova In a fairy tale
When I think of you	oh	Casanova I feel love would
feel it in your heart?	oh	Casanova In a fairy tale
dream this dream with you	oh	Casanova I feel love would
feel it in your heart?	oh	Casanova In a fairy tale
can't live if we're apart	oh	Casanova Don't come down from
be Only a memory remains	oh	yes your heart belongs to
be Only a memory remains	oh	yes your heart belongs to
be Only a memory remains	oh	yes your heart belongs to

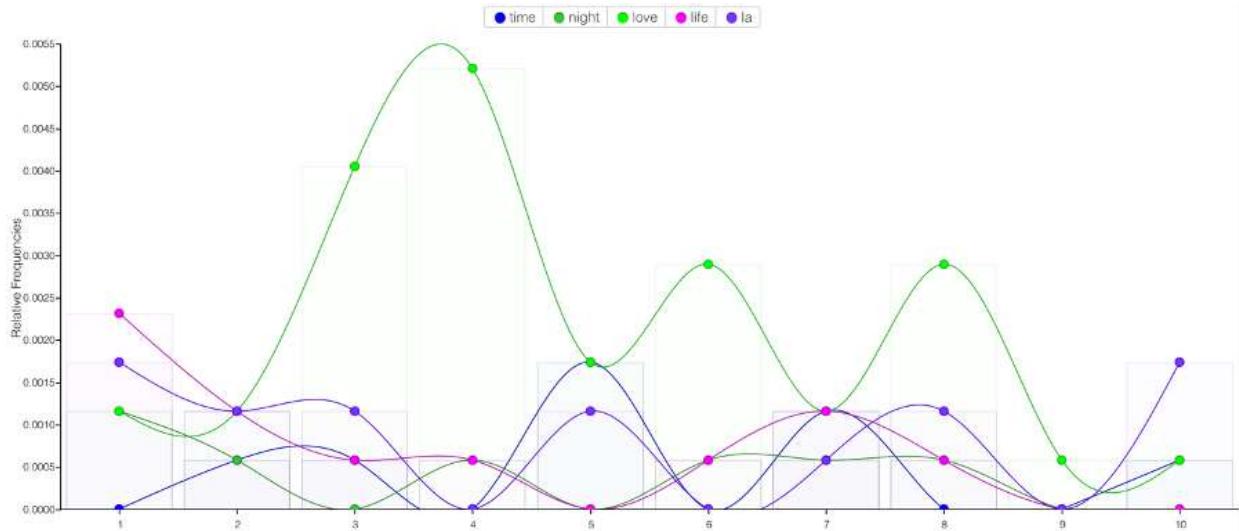
Left	Term	Right
when I'm around you I	can't	live if we're apart I
when I'm around you I	can't	live if we're apart You
when I'm around you I	can't	live if we're apart Oh
our future is standing there	can't	you see you're my future
So don't you pretend you	can't	get it why I say
Though you're gone forever I	can't	make you disappear Crashing breaking
Though you're gone forever I	can't	make you disappear Cause you
heart of stone And I	can't	get used to being all
Though you're gone forever I	can't	make you disappear No no
overwhelming me You're too perfect,	can't	you know? Please, get out
ain't key to me I	can't	think that easily If you
Without me you cannot be	can't	ya see, we aren't free
you distract luck But you	can't	deny it's a mistake Leaving
am I gonna do Ca-	can't	deny it, deny it, what
but I'm telling I just	can't	believe it Never thought that
but I'm telling I just	can't	believe it Never thought that

Topic analysis also reinforces this pattern - words such as *love*, *life* and *time* dominate both Cirrus and Terms, with *love* appearing 37 times.



Term	Count
love	37
la	15
life	12
time	8
right	7
je	7
heart	7
dance	6
song	5
day	5
night	5
lights	5
free	5
feel	5
storm	4
running	4
run	4
oh	4
kiss	4

Trends indicate that these themes remain consistently present throughout the timeframe.

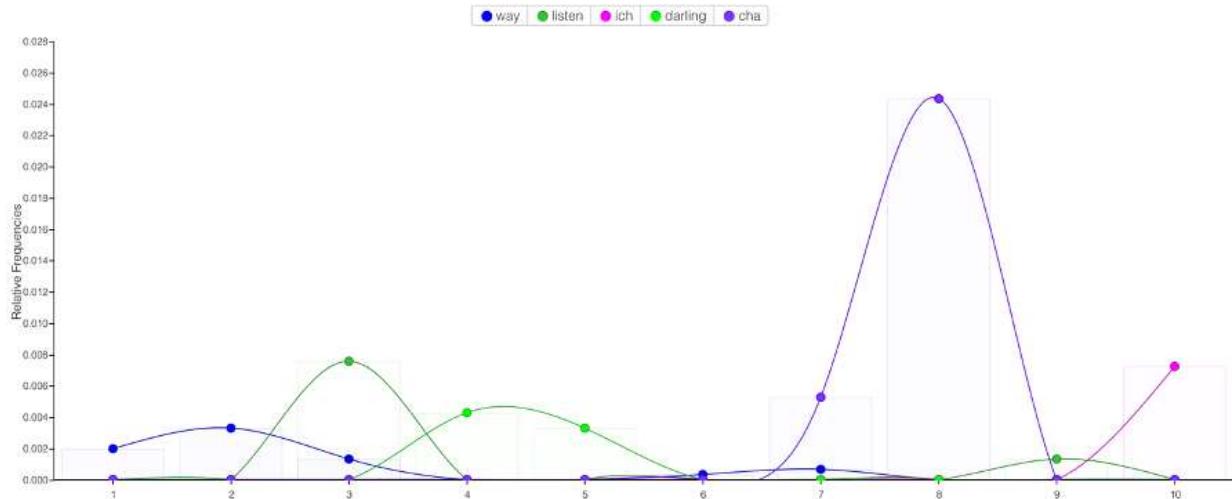


The overlap between lyrical repetition and thematic stability suggests that Indo-European songs strongly align with Eurovision's stylistic and conceptual norms.

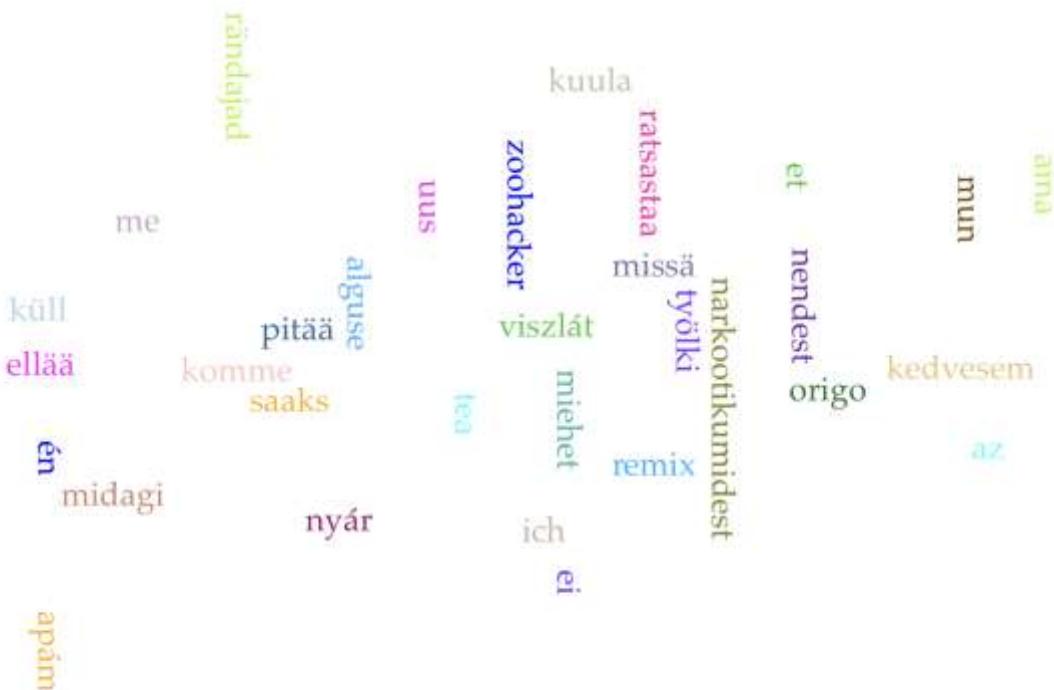
## Uralic languages: lyrical flexibility and thematic distinction

Songs in languages from the Uralic family (i.e. Estonian, Finnish, Sami, Udmurt and Hungarian) rely less on repetition and show greater thematic variation.





The Uralic language group presents a contrasting pattern to the other language families. In the lyrical analysis, Cirrus highlights sound-based elements such as *cha*, which appear frequently (90 times) but carry little semantic weight, being used in just one song. Terms and Contexts show lower repetition overall, and Trends reveal greater fluctuation across years. However, topic analysis tells a different story.



Although the topic corpus is smaller (33 topic words), the dominant terms are semantically meaningful and less standardized than in the Indo-European group. Contexts show that these words are used in more specific narrative environments.

Left	Term	Right
Viszlát nyár Az én apám	cha	Cha Cha (nendest) narkootikumidest ei
nyár Az én apám Cha	cha	Cha (nendest) narkootikumidest ei tea
Az én apám Cha Cha	cha	(nendest) narkootikumidest ei tea me
Left	Term	Right
saaks alguse Kedvesem (Zoothacker Remix)	aina	mun pitää Origo Viszlát nyár

This divergence suggests that Uralic songs may conform rhythmically to Eurovision norms while maintaining greater thematic individuality.

**Other language families (Afro-Asiatic, Kartvelian, Turkic, as well as imaginary languages): diversity and analytical limits**

Smaller language groups resist standardization, emphasizing diversity over repetition.



Left	Term	Right
Sheet1 O julissi	na	jalyni O julissi na dytini
julissi na jalyni O julissi	na	dytini O bulo diti non
Sestrone dina katsu O julissi	na	ti buku O julissi na
na ti buku O julissi	na	katinu Dvoranu mojani bidna Marusi
marusinja Kolosali krokodili O julissi	na	jalyni O julissi na dytini
julissi na jalyni O julissi	na	dytini O bulo diti non
Sestrone dina katsu O julissi	na	ti buku O julissi na
na ti buku O julissi	na	katinu Dvoranu mojani bidna Marusi
marusinja Kolosali krokodili O julissi	na	jalyni O julissi na dytini
julissi na jalyni O julissi	na	dytini O bulo diti non
Sestrone dina katsu O julissi	na	slukati O julissi na kotchali
julissi na slukati O julissi	na	kotchali Od nu je dvorian
sti budo najali O julissi	na	ja O julissi na jalyni
julissi na ja O julissi	na	jalyni O julissi na dytini
julissi na jalyni O julissi	na	dytini O bulo diti non
Sestrone dina katsu O julissi	na	jalyni O julissi na dytini
julissi na jalyni O julissi	na	dytini O bulo diti non

The group of other language families displays the greatest diversity. In lyrical analysis, Cirrus highlights phonetic elements such as *na*, while Terms show a fragmented vocabulary with no dominant core. Contexts confirm that many frequent elements function as sound rather than meaning.

With only 24 topic words, no single theme dominates and the Trends tool shows irregular usage across years. This group demonstrates the limits of frequency-based comparison and highlights how some Eurovision entries prioritize linguistic identity over shared thematic conventions.

### Conclusion: shared form, divergent meaning

The comparison of lyrics and topics across language families reveals a consistent pattern: while Eurovision songs tend to converge in lyrical form, they diverge more clearly in thematic focus. Across all groups, lyrics rely heavily on a small set of high-frequency words and sound-based elements, reflecting the performative and musical demands of the Contest. These patterns remain stable over time, suggesting that repetition and simplicity are structural features of Eurovision songwriting.

Topic-based analysis, however, exposes meaningful differences between language families. Indo-European songs show strong thematic standardization, while Uralic and other language families display greater variation and specificity. In these groups, themes are less constrained by dominant Eurovision conventions and more closely tied to linguistic or cultural

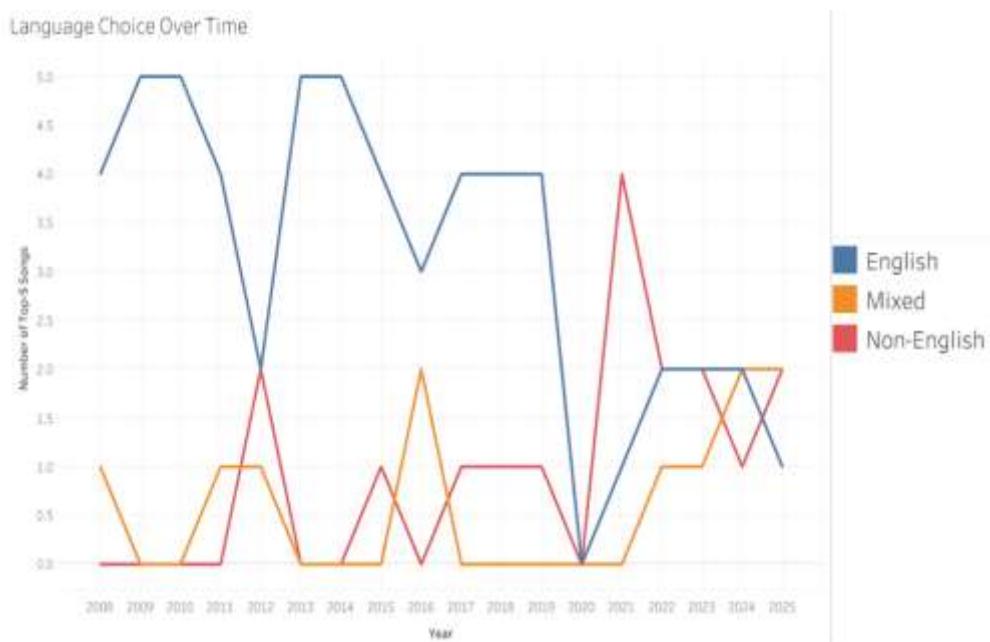
contexts. This contrast highlights how shared musical forms can coexist with distinct approaches to meaning.

Overall, the findings demonstrate that language family plays an important role in shaping how themes are expressed, even within a highly standardized musical environment. Our hypothesis about same language family songs sharing greater lexical and thematic similarity than songs across different families is therefore generally confirmed. By combining multiple Voyant tools, this project shows how digital text analysis can uncover both common structures and subtle differences, offering a clearer understanding of linguistic diversity within Eurovision.

## **TEXT ANALYSIS ACROSS TOP 5 SONGS**

Moving on to a related text analysis based sub-question, we seek to explore: How have the language choice and topic/vocabulary trends of each contest's top five most highly rated songs change between 2008 and 2025? This analysis is based on data from the top five ranked Eurovision Song Contest entries each year (excluding 2020, when the Contest was cancelled with no voting), using Voyant Tools and Tableau Public for visualizing our text analysis. By focusing only on the highest ranked songs, we aim to highlight the linguistic and lyrical patterns that are linked with success in the contest.

### **Language choice over time**



This visualization clearly shows that English has been the dominant language among the top five Eurovision songs throughout the 2008-2025 period, particularly up until 2020. In most of these years, the majority of highly ranked songs have been performed fully in English, an exception being 2012. This suggests that English has generally been viewed as the safest and most effective language choice for reaching a wider international audience and securing more points.

However, the chart also reveals that in recent years, since 2020, more highly ranked participants choose to perform mixed language songs where another language is included. Moreover, in 2021, 4 out of 5 highly ranked songs weren't performed in English at all, with countries like France, Ukraine, Switzerland and the winner Italy choosing to perform in their native languages, perhaps influenced by the cultural isolation and promotion of local culture relevant in many countries during the preceding COVID-19 pandemic. In the following years, it is also evident that more participants choose to perform in their national language or mix it with another language with English. This indicates that while English is still commonly used, it is no longer viewed as almost a requirement for reaching a high result.

## Vocabulary and topic changes

When analyzing how lyrics and vocabulary change over time, there were some noticeable patterns. In the earlier years of the timeframe, especially between 2008 and 2015, the most common words related to romantic relationships. Vocabulary such as *love*, *like*, *heart*, *dream*, *beautiful* and *believe* appears frequently. This reflects a common assumption about the Eurovision “style” of music - the general public opinion that Eurovision songs often focus on love, longing and emotional connection, typically between two people.



Over time, the vocabulary trends can be seen to change. From the mid-2010s onwards, there is a noticeable increase in self-focused language, with words such as *you*, *I*, *me* and *we* becoming more common. Many songs focus on the artists' own feelings, experiences or personal struggles. This shift showcases a move away from classic love songs and towards more individually reflective songs. However, the romantic theme doesn't disappear - love-related words continue to appear frequently throughout this period. Additionally, action-driven words (*come*, *make*, *let's*) remain popular. This shows how familiar themes are adapted to match changing cultural values, where nowadays the focus in music is often placed on themes of acceptance, self-love and emotional strength and growth.



## Overall trends

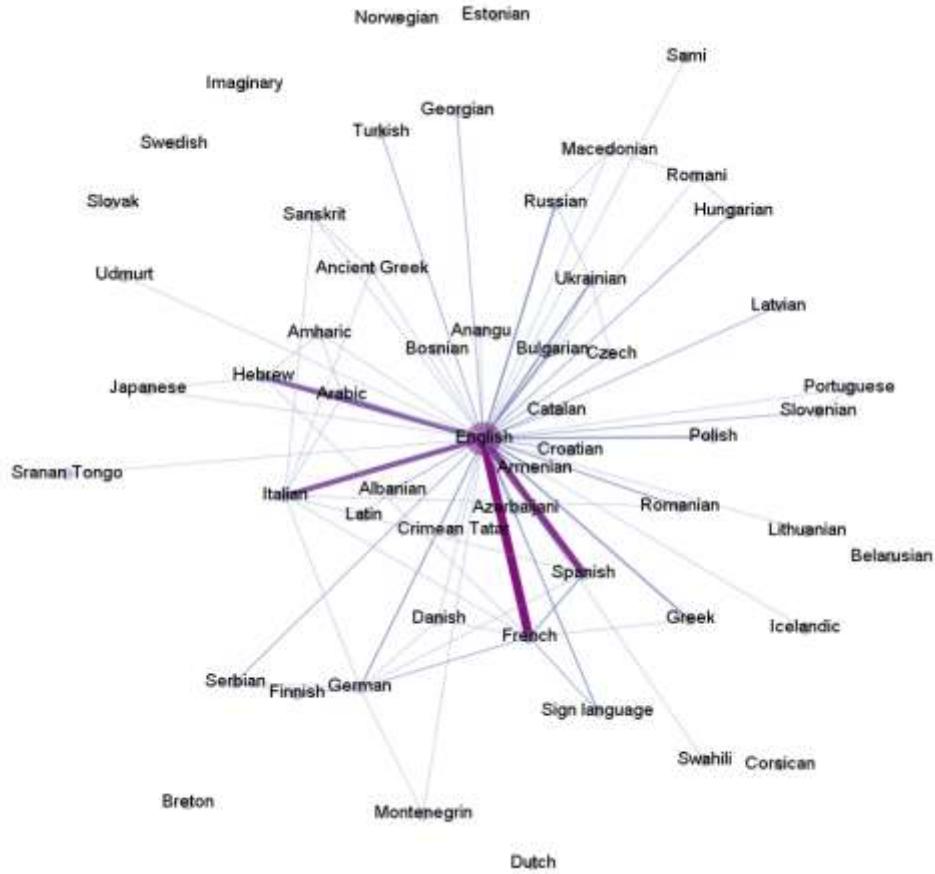
When viewed together, the language and vocabulary trends suggest that Eurovision's top 5 songs every year increasingly emphasize authenticity and self-expression. Earlier entries appear to aim for universal appeal through English lyrics and love stories, while later entries are more daring to experiment with identity and action-driven thematic focuses and language choices. The growing acceptance of non-English songs among the top five supports this suggestion and somewhat disproves our hypothesis of highly ranked songs displaying more standardized language choices and thematic patterns over time. Instead of tying success to the use of the English language, the use of the national language can strengthen a song's emotional impact and national identity, allowing it to be less standardized and more personal, while still being relatable for certain audience groups. This suggests that Eurovision has moved away from one well-known formula for success and now has more space for rewarding originality, diversity and confidence.

Years with fewer fully English language songs among the top five challenge the idea that English is always the best choice. These cases often involve strong performances that rely on clear emotions, a narrative or distinctive musical style. The data suggests that when these elements are present, a non-English language becomes less of a barrier and still receives recognition from a global audience.

Overall, the data shows that between 2008 and 2025, the top five Eurovision songs generally have become more linguistically diverse and thematically complex. English still remains as the most common language, but its dominance has weakened over recent years. At the same time, lyrical topics have shifted from traditional romantic narrative towards themes of identity, personal growth and empowerment. The visualizations created by Tableau and Voyant tools make the changes easier to compare across years.

## **SOCIAL NETWORK ANALYSIS**

In this section, we use social network analysis to investigate the sub-research question: How are different languages used together in bilingual or multilingual songs, particularly with English functioning as a lingua franca on the international stage? The Gephi visualization software allows us to create a language network where the nodes are all 55 languages used in the 2008-2025 Eurovision Song Contests, and the weighted edges represent the number of times each connected language pair has been used together in a song, showcasing both dominant co-use and meaningful outliers.



The most noticeable pattern is the centrality of English, which is by far the most connected language in the network, acting as a hub through which the majority of other languages are linked. The strongest edges in the dataset all involve English, with English-French (16 times), English-Spanish (13), English-Italian (10) and English-Hebrew (10) co-usage standing out the most. This visually affirms English language's role as a lingua franca (common shared language among speakers of different native languages) on the international music stage, and also partially confirms our hypothesis that English will most often appear alongside national languages in multilingual songs as a means of increasing accessibility, rather than replacing local identity. However, in most bilingual songs English takes the dominant role, with the majority of lyrics being in English and only select words/phrases being in the other, typically native language of the artist, suggesting that this accessibility may be prioritized over more authentic expressions of local identity.

Interestingly, the English-French connections are not solely driven by artists from France, as only seven of the sixteen song examples originate there. This suggests that French, more than any other language, is used as a stylistic choice beyond national boundaries, often paired with

English to evoke cultural or genre-specific associations. This also makes sense considering that historically, French has been a more local shared language for many European countries where it is an official language, and that French, along with English, is the joint official language of the European Broadcasting Union, with the Contest's hosts always presenting information in both languages as a tradition.

The dominance of Romance languages (French, Spanish and Italian) in the strongest English pairings reveals a narrative of linguistic compatibility shaped by global pop traditions. Romance languages are widely recognizable and melodic when used in popular music genres (e.g. Latin music), making them complementary to English. At the same time, the presence of English-Hebrew as an equally strong pairing complicates a purely Eurocentric reading. Belonging to the Afro-Asiatic language family's Semitic language branch, the presence of Hebrew language suggests that diaspora audiences and historic geopolitical connections to the contest can also influence multilingual practices in Eurovision. Thus, while linguistic and geographic proximity between countries is seen strongest, it is not the only factor affecting language co-use.

Beyond the strongest Romance language edges, English also connects moderately well to other Indo-European language branches, especially Slavic and Hellenic, as represented by Ukrainian, Russian, Polish, Bulgarian and Greek (each linked 4 times). These connections indicate a pattern in which English serves as a bridge language, enabling artists from smaller linguistic markets to participate in the international music industry. The visualization also reveals differences in connectivity – many languages appear only in conjunction with English and never connect to each other directly. This pattern suggests that multilingualism in Eurovision is often asymmetrical: the first and strongest bilingual pairing not including English, which is the French-Spanish pair, only appears 3 times, emphasizing how rarely multilingual songs exclude English entirely. This further reinforces English's gatekeeping function in Eurovision, especially at the beginning of our timeframe.

Looking beyond the most visible patterns, the nodes of unusual languages such as Latin and Ancient Greek may represent symbolic rather than communicative linguistic functions. These languages are not used for everyday speaking, but rather artistic purposes, often symbolizing historic or intellectual topics. Another important narrative concerns the representation of artist heritage: languages such as Sranan Tongo, Sami, Swahili, Sanskrit and Arabic appear as peripheral nodes, often connected only once, typically to English. They suggest that multilingual songs are

also spaces for expressing more personal cultural identities that may differ from the representative country. English in these cases can act as a bridge that allows languages from beyond Eurovision-participating countries to be perceived within music by the international audience. Finally, nine outliers – Belarusian, Breton, Corsican, Dutch, Estonian, Norwegian, Slovak, Swedish and even imaginary language – appear as isolated nodes. These outliers appear to reflect use of language to represent local ethnic groups (e.g. Breton and Corsican in France) or to express purely local identities, relying on the strength of the message within music rather than more accessible lyrics in English.

Overall, the network visualization demonstrates that bilingual and multilingual Eurovision songs are shaped by a balance between international reach and local identity. The English language tends to take on the role of a bridge for the audience members accessing international music beyond the boundaries of their native languages, confirming our initial hypothesis, while exposure to other languages contributes cultural identity representation and narrative depth to the songs.

## **MAP ANALYSIS**

In this section, we use geospatial visualization and analysis to investigate the sub-research question: How have participating countries used other nations' languages in their songs? Since language use does not conform to a single country's borders, GIS visualization via the QGIS platform is a helpful way to spatially represent language use patterns in the shared cultural space of Eurovision, depicting the relationships of 48 participating countries and their 43 official languages.



*(Interactive map is accessible on the [project website](#))*

This map visualizes the relationships between a source country and target countries whose official languages the selected source country has used in its songs, with the labels providing information on the number of such songs and the years they were sent to the Eurovision Song Contest. Between 2008 and 2025, some Eurovision participants have increasingly integrated languages from other nations into their performances, although the priority, especially in bilingual native-English language pairings, continues to be the country's own official or minority language. The use of English is present across nearly all participating countries, corresponding to the United Kingdom, Ireland, Malta and Australia as target countries.

Besides these countries themselves, only 2 countries have used English in every song from this period – Azerbaijan and Germany (all 18 times), followed by Belgium, Denmark, Norway and Sweden (17 times) and Latvia (16 times). On Germany's part, despite German being a globally significant language and Germany being one of the Big 5 countries with automatic entry into the Eurovision final each year, German language songs have historically performed inconsistently at Eurovision in this timeframe. Since 2008, the local broadcaster has selected artists already producing English language music for global markets, seemingly having less faith in the success of their own language based on historic voting trends. Similarly, in the case of Azerbaijan (and countries like Sweden and Denmark), an explanation could be that historic success when sending English language songs has reinforced English as the “safe” option. The presence of 3 Scandinavian countries in this statistic appears to reflect the high English proficiency and the local music industry's West-leaning nature, while smaller countries like Belgium and Latvia may choose to rely on English as a language with significantly larger reach than their own small populations' languages.

Besides English, geospatial patterns of regional linguistic borrowing also appear, like Finland having targeted use of Swedish (a minority language in Finland) in 2012 and German in 2025, while Latvia and Lithuania each feature songs in Russian in 2009. These patterns reveal cross-border cultural proximity, and, in the case of the Baltic countries – the presence of Russian-speaking minorities, the post-Soviet cultural influences and the strategic choice based on the Contest being held in Moscow that year. Similarly, smaller countries such as Luxembourg, San Marino and Malta occasionally incorporate languages also used by their larger neighbors, reflecting both historical linguistic influences and practical considerations of audience reach. For example, San Marino incorporates both Italian and Spanish across different years, seemingly using

a linguistic strategy to gain listeners from neighboring countries beyond its limited domestic audience.

Looking at countries with less diverse language choices, the native language is most favored by Spain (17 times), France (16), Italy and Portugal (15) and Serbia (14). All except Serbian are Romance languages from Southern Europe, reflecting these languages' associations with emotional expressiveness, established international music presence (e.g. Portuguese and Spanish in South America). Three of these countries (Spain, France and Italy) are also part of the Big 5, so the guaranteed entry to the Eurovision final reduces pressure to rely on English as a lingua franca for audience reach. In the case of Serbia, its strong self-targeting contrasts with neighboring Balkan countries that rely more heavily on English, suggesting that Serbia's broadcaster might have consciously chosen a national identity-driven linguistic approach to Eurovision as an international stage. Similarly, geopolitical factors appear to have influenced a shift in Ukraine's language choices: earlier entries were almost exclusively in English, but every entry since 2020 has been in Ukrainian or joint English-Ukrainian, reflecting the need for increased local cultural identity expression due to the pressures of the ongoing Russo-Ukrainian war.

Overall, the map suggests two parallel narratives: the reliance on English as a bridging language shared by speakers across Europe, as well as the culturally motivated borrowing of regional languages, reflecting historical ties, identity expression and audience targeting. The hypothesis that the use of non-native languages will reflect cultural affiliation or regional proximity between countries is confirmed by the examples above, also aligning with the analysis of the social network visualization in the previous section.

## **CONCLUSIONS**

Our project set out to examine how Eurovision Song Contest participants' linguistic choices between 2008 and 2025 reflect broader geopolitical, cultural and communicative dynamics. Across vocabulary and topic analysis, social network analysis and geospatial visualizations, a shared conclusion can be made: Eurovision acts as a shared cultural space not only for the content of the presented songs, but also for the languages and vocabulary they contain. Both linguistic standardization and diversity coexist here, shaped by the countries' geopolitical and cultural ties as well as the representing artists' personal, identity-driven expression.

Textual analysis across language families demonstrates that while Eurovision lyrics converge strongly in form, relying on a shared, repetitive performative vocabulary, they diverge more clearly at the level of thematic focus. Language families influence how meaning is structured, with Indo-European languages aligning most closely with Eurovision's dominant norms, while Uralic and smaller language groups retain greater thematic individuality.

Analysis of the top five ranked songs reveals that, although English has historically functioned as a dominant and “safe” language choice, its dominance has lessened in recent years, with non-English and multilingual entries increasingly achieving high rankings. This shift suggests growing audience acceptance of linguistic authenticity and cultural specificity.

Social network analysis further confirms English language’s role as a lingua franca, acting as a central bridge language in bilingual and multilingual songs, while also revealing stylistic and symbolic uses of other languages that extend beyond national borders. Finally, the geospatial language map highlights how non-native language use often follows patterns of regional proximity, shared linguistic heritage or strategic audience targeting, while some countries choose to use their own languages due to their reach in the global music industry or in order to highlight their national identity on the international stage.

Altogether, these findings show that Eurovision’s linguistic landscape reflects a negotiation between international accessibility and local identity, revealing how language functions as both a communicative tool and a cultural symbol within Europe’s most visible musical event.

## **BIBLIOGRAPHY**

Adamska, K., & Reiss, J. (2025). Predicting Eurovision Song Contest results: A hit song science approach. *Transactions of the International Society for Music Information Retrieval*, 8(1), 93–107. <https://transactions.ismir.net/articles/10.5334/tismir.214>

Ahlberg, M. (2023). Beyond the music: Exploring the dynamics of national and European identity in the Eurovision Song Contest (Publication ID - URN: urn:nbn:se:mau:diva-62148) [Bachelor’s thesis, Malmö University]. Digitala Vetenskapliga Arkivet.

Barbeau, E. (2022). *English language dominance in the Eurovision Song Contest: A sociolinguistic analysis* [Unpublished master’s thesis]. Vytautas Magnus University.

<https://portalcris.vdu.lt/server/api/core/bitstreams/f3c60b99-ff5e-435a940c88d53124680c/content>

Bohlman, A. F., & Polychronakis, I. (2013). Eurovision everywhere: A kaleidoscopic vision of the Grand Prix. In D. Tragaki (Ed.), *Empire of song: Europe and nation in the Eurovision Song Contest* (pp. 57-73). Scarecrow Press.

Comrie, B. (Ed.) (2009). *The world's major languages* (2nd ed.). Routledge.  
<https://doi.org/10.4324/9780203301524>

minitree. (2022). *Lyrics of every Eurovision song in the history* [Data set]. Kaggle.  
<https://www.kaggle.com/datasets/minitree/eurovision-song-lyrics>

Motschenbacher, H. (2016). Language choice practices in the ESC. In *Language, normativity and Europeanisation. Postdisciplinary studies in discourse* (pp. 117–151). Palgrave Macmillan, London. [https://doi.org/10.1057/978-1-137-56301-9\\_4](https://doi.org/10.1057/978-1-137-56301-9_4)

Palacios-Hidalgo, F. J., & Huertas-Abril, C. A. (2025). Do songs in English do better? Studying linguistic diversity within Eurovision Song Contest entries. *Diacritica*, 39(2), 389–403.  
<https://doi.org/10.21814/diacritica.6280>

Spijkervet, J. (2020). *The Eurovision dataset* (Version 3.0) [Data set]. Zenodo.  
<https://doi.org/10.5281/zenodo.10088892>

Thorne, S. L., & Ivković, D. (2015). Multilingual Eurovision meets plurilingual YouTube: Linguascaping discursive ontologies. In D. A. Koike & C. S. Blyth (Eds.), *Dialogue in multilingual and multimodal communities* (pp. 167-192). John Benjamins Publishing Company.

Torres, G. (2022). Pan-European, national, regional and minority identities in the Eurovision Song Contest. In K. Jakubowicz & M. Sükösd (Eds.), *Media, nationalism and European identities* (pp. 247-268). Central European University Press.