Fast & Furious: Semantic Drift

Patrik Stano

NLP Centre, Faculty of Informatics, Masaryk University
Botanická 68a, Brno, Czech Republic
nlp.fi.muni.cz/en

Abstract. This paper examines the evolving semantics of family across the Fast & Furious film series (2001–2023) through computational linguistic analysis. Using the framework of SemAxis, we construct a semantic dimension spanning kinship and ideological conceptions of family. Sentence-level embeddings from movie subtitles are projected onto this axis to quantify how the term's meaning changes between and within individual films. The semantic projection of the original English subtitles is then compared to a Czech translation to examine how this change affects the results. Complementary measures, including lexical collocate drift, frequency analysis, and intra-movie ideological trajectories, reveal a systematic transition from literal familial relations toward a loyalty-based ideology. The results demonstrate a measurable stabilization of ideological semantics over time: early films show high intra-narrative variability, while later entries converge on a narrow, consistent interpretation. This study illustrates how long-form media franchises can encode and reinforce cultural redefinitions of social concepts through language.

Keywords: Semantic axes; Embeddings; Fast & Furious

1 Introduction

Across more than two decades, the Fast & Furious franchise by [4] has become synonymous with a single word: family. The frequent, and increasing, use of this term sparked interest even among those who would not be considered the target audience of the action-packed and car-focused movies. Yet the term's usage has evolved dramatically—from literal references to parents, siblings, and domestic bonds in early installments to an abstract, almost ritualized ideology of loyalty and belonging in later ones, where family can be viewed as an expression of the moral code of a heroic team. This linguistic transformation provides a compelling case study in semantic drift, the gradual change in meaning that accompanies cultural and narrative evolution.

This study aims to quantify this transformation by tracking how family is used in all ten mainline films released so far (2001–2023). Our analysis combines sentence embeddings and cosine similarity metrics to model a semantic axis between "kinship" and "ideological" interpretations of family, inspired by the SemAxis approach proposed by [1]. We complement this with measures of

A. Horák, P. Rychlý, A. Rambousek (eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2025, pp. 109–116, 2025.

lexical drift in collocate distributions, frequency trends, and an analysis of intra-movie ideological trajectories, offering a comprehensive linguistic portrait of how family becomes a moral principle rather than a literal concept.

2 Methodology

2.1 Data

The corpus used consists of the full subtitle transcripts from the ten mainline Fast & Furious films released between 2001 and 2023. The subtitles were collected from publicly available .srt files and processed using a preprocessing pipeline. The cleaning procedure removed timestamps, speaker names, and non-verbal annotations using regular expressions while preserving spoken dialog, ensuring that the linguistic material analyzed reflects the natural language of the films rather than metadata or sound cues. For the same reason, song lyrics were removed from the subtitle files, since these could skew the semantic analysis, as their content often does not directly correlate with the plot of the films.

Each subtitle file was converted to plain text and split into sentences using NLTK [2]. The resulting dataset contained 8218 sentences in all films, with individual movie counts ranging from 540 to 1080, depending on script length and dialog density.

The cleaned dataset was stored in a structured csv format , which served as the primary input for the semantic modeling pipeline described below.

2.2 Semantic Axis Construction

To model the conceptual drift of family, we adopt the SemAxis framework introduced by [1], which defines semantic meaning as a position in a continuous vector space spanned by two opposing sets of seed terms. In our adaptation, the poles of the axis correspond to two contrasting interpretations of family: a kinship-based sense grounded in domestic and biological relations and an ideological sense emphasizing loyalty, unity, and belonging. Two curated lists of 20 seed phrases were designed to capture these extremes. The kinship seeds included expressions such as ″my mother and father″ and ″our parents raised us″, while the ideology seeds comprised phrases like ″we protect our own″ and ″loyalty means everything″. These seed sets were deliberately chosen as phrases rather than single words to better capture the contexts as well as fitting a sentence transformer directly.

Each phrase was embedded using the SentenceTransformer model *sentence-transformers/paraphrase-multilingual-mpnet-base-v2*, introduced by [3], chosen for its popularity in semantics-based tasks as well as its multilingual capabilities needed for the Czech comparison. The mean vector of each seed set was computed to form prototype embeddings for the kinship and ideological poles. For each sentence in the dataset, cosine similarity was measured relative to both prototypes, and the resulting ideology index was defined as:

$$\textit{Ideology Index} = \cos(\text{sentence}, \text{ideology}) - \cos(\text{sentence}, \text{kinship}) \tag{1}$$

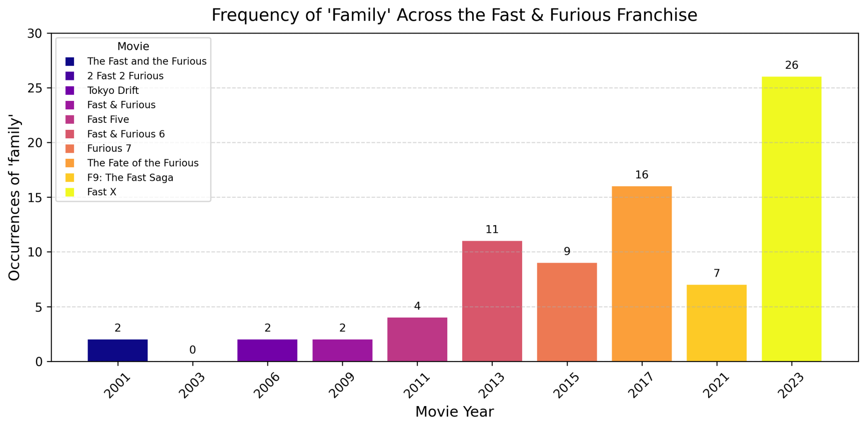

Fig. 1: Occurences of the word ”family” across the movies

A positive index value indicates alignment with ideological semantics (loyalty, unity, belief), whereas a negative value indicates kinship-oriented language (parents, siblings, blood relatives).

2.3 Additional measures

To complement this projection-based analysis, three additional computational measures were implemented:

- **Frequency Analysis** : the raw count of the word ”family” was computed for each movie as a part of data exploration.
- **Lexical Semantic Drift** : for each film, a distribution of the collocates of family was extracted, and cosine similarity between consecutive distributions was calculated. This quantifies lexical drift, capturing how the immediate linguistic context of family changes across the series.
- **Within-Movie Shifts** : to trace how ideology evolves within individual films, each movie's sentences were divided into ten equal-length narrative bins. The mean ideology index per bin was plotted to visualize internal trajectories, revealing differences in intra-narrative variation between early and late installments. Finally, the magnitude of variation in these within-movie trends was quantified by calculating the area between the highest and lowest ideological trajectories across the series halves (2001–2011 vs. 2013–2023). This metric showed a marked contraction of variability, indicating semantic stabilization in later films.

3 Results

3.1 Frequency Analysis

The lexical frequency of the term family provides a straightforward indicator of its salience across the franchise. As shown in Figure 1, the number of family

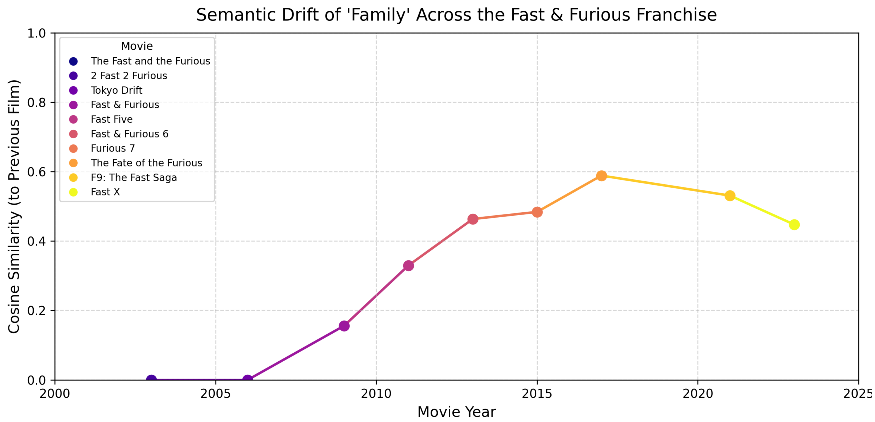

Fig. 2: Semantic drift of family collocates across the franchise

mentions increases markedly over time, rising from a few instances in the first films to nearly thirty in Fast X, meaning one mention of family per approximately 6 minutes of film. This upward trend mirrors the growing centrality of family in the films' discourse and marketing, transforming it from a background motif into a defining theme. However, frequency alone cannot distinguish between literal and figurative uses. This motivates a more fine-grained semantic analysis of contextual meaning, developed in the following subsections.

3.2 Lexical Semantic Drift

To examine how the contextual meaning of family evolves, collocate distributions were computed for each movie and compared using cosine similarity. Figure 2 presents the resulting semantic drift curve, where lower cosine similarity values indicate larger changes in the lexical neighborhood of family. The first two points are equal to zero since the word does not appear in the script of 2 Fast 2 Furious. The trajectory reveals two distinct sections: substantial contextual change in the early series (2001–2013), followed by stabilization from Furious 7 (2015) onward. This lexical convergence suggests that the meaning of family became increasingly standardized, possibly in the recurring ideological formula of loyalty, unity, and shared mission rather than fluctuating according to current character relationships. However, discovering what it is stabilizing towards requires further research, which is explored in the following section.

3.3 Cross-Movie Ideology Trend

Projecting each sentence onto the kinship–ideology axis yields a quantitative measure of ideological alignment. As shown in Figure 3, the mean ideology index, calculated as a mean of the value of each sentence in a film, rises steadily across the franchise, with a statistically significant linear trend (slope=0.0014, $p=0.0129$, $R^2=0.559$). The inclusion of the 95 % confidence interval confirms that

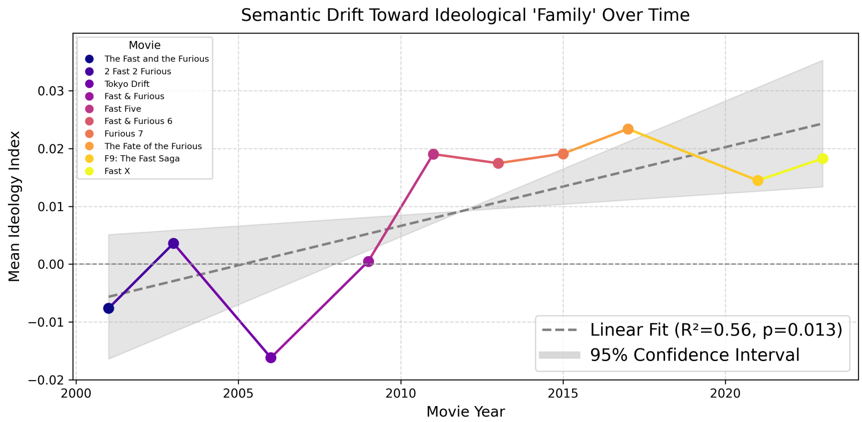

Fig. 3: Ideology index shift across the franchise (English)

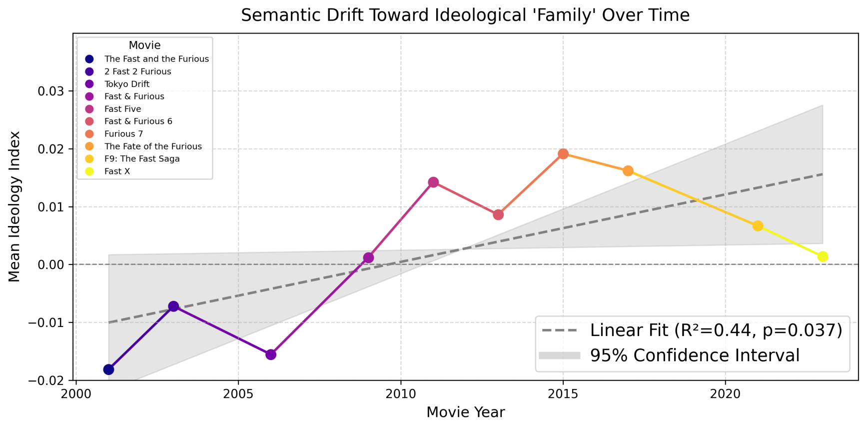

Fig. 4: Ideology index shift across the franchise (Czech)

this progression is unlikely to be random. This indicates that linguistic contexts surrounding family have drifted from literal kinship semantics toward collective ideology, mirroring the franchise's thematic evolution from local street-racing affairs to global, quasi-mythic brotherhood. Compared to the Czech subtitle version shown in Figure 4, we see that the trend is weaker, indicating that a part of the director's vision was lost in translation.

3.4 Within-Movie Shifts

The within-movie analysis shown in Figure 5, Figure 6, and Figure 7 divides each film into ten equal narrative segments and computes the mean ideology index per segment. The shift observed in the previous section repeats here, as earlier films, shown in hues of blue, appear lower overall compared to later installments, shown in orange and yellow. Early installments display considerable internal variation, with ideological tone rising and falling throughout the plot. Later films, by contrast, exhibit smoother and more uniformly ideological pro-

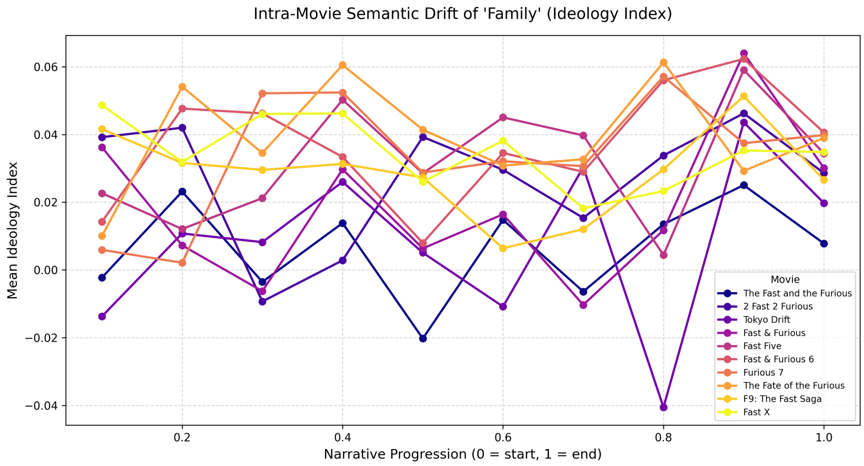

Fig. 5: Shift of the ideology index within movies

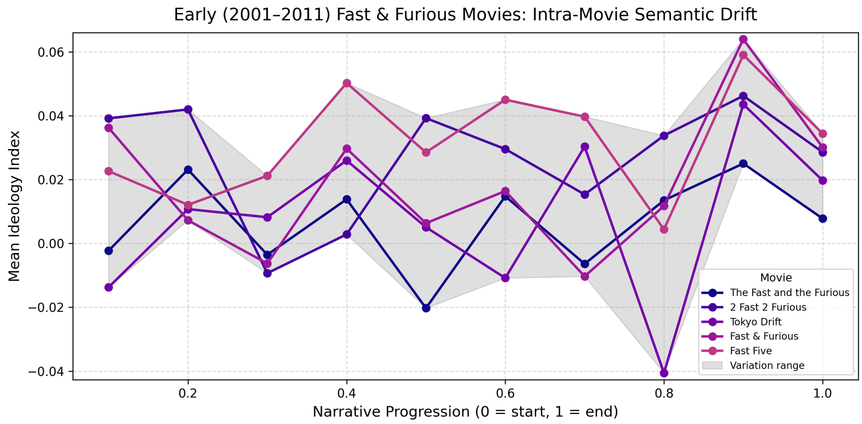

Fig. 6: Ideology index 2001-2011

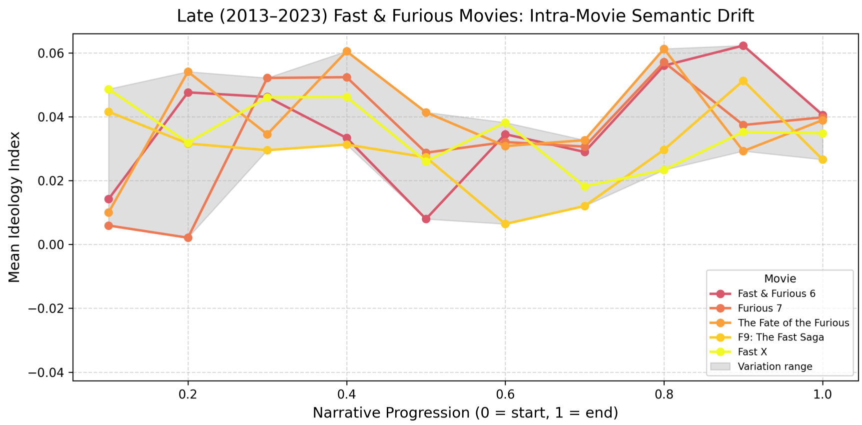

Fig. 7: Ideology index 2013-2023

files, indicating semantic stabilization both within and across movies. To illustrate this contrast, two aggregated subplots were generated: one for the first five films (2001–2011) and one for the later five (2013–2023). Both use identical y-axis scaling to enable direct comparison. The area between ideological trajectories, a measure of semantic variability, contracts by 32.95% in the latter half, confirming that the concept of family became more consistent codified over time.

4 Discussion

The analysis presented above demonstrate a clear semantic reorientation of the concept family across the Fast & Furious film series. Although the earliest films deploy family in its literal, kinship-oriented sense, later installments redefine it as a marker of ideological affiliation: loyalty, unity, and collective identity. This linguistic shift parallels the narrative expansion of the franchise from small-scale street racing dramas to global action epics centered on moral codes and symbolic brotherhood, which persists even after the movies translation into Czech.

From a computational perspective, the semantic axis approach derived from SemAxis [1] provides an interpretable representation of this drift. By projecting entire sentences into a vector space defined by kinship and ideological prototypes, it becomes possible to trace and quantify not only the frequency of thematic references but also their semantic orientation. The lexical drift results reinforce this interpretation: early variability in collocate distributions corresponds to flexible and context-sensitive usage, whereas later films converge on a fixed formula. Similarly, within-movie analyses reveal that semantic variation decreases over time, suggesting that the films' internal structure became more homogeneous. This stabilization mirrors broader industrial patterns of franchise storytelling, where repetition and brand coherence replace experimentation. Importantly, these trends also reflect Vin Diesel's growing creative control beginning with Tokyo Drift (2006), after which the recurring invocation of family became a narrative motif. The word's semantic narrowing therefore corresponds not only to character development but to the institutional consolidation of the franchise's cultural identity.

5 Conclusion

In broader terms, this study illustrates how computational semantics can reveal ideological evolution within long-form media franchises. By quantifying the change of meaning through distributional and projection-based methods, we can measure shifts that align with narrative and authorship transformation. Here, linguistic evidence supports the cultural intuition that Fast & Furious has turned family into a moral code—less about kinship, more about belonging and loyalty in an imagined global community.

References

1. An, J., Kwak, H., Ahn, Y.Y.: SemAxis: A lightweight framework to characterize domain-specific word semantics beyond sentiment. In: Gurevych, I., Miyao, Y. (eds.) Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 2450–2461. Association for Computational Linguistics, Melbourne, Australia (Jul 2018). https://doi.org/10.18653/v1/P18-1228, https://aclanthology.org/P18-1228/
2. Loper, E., Bird, S.: Nltk: The natural language toolkit (2002), https://arxiv.org/abs/cs/0205028
3. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics (2019), https://arxiv.org/abs/1908.10084
4. Universal Pictures: Fast & furious [film franchise]. Film series (2001), ongoing film series, 2001–present. Directed by various directors including Rob Cohen, Justin Lin, and James Wan. Universal Pictures.