Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

DATA REPORT JUNE 26, 2026 UPDATED AUGUST 6, 2026

We Counted 47,534 Words of Dialogue From 16 Iconic Jewish Screen Characters. What Does Their Language Reveal?

A lexical analysis of 47,534 words of dialogue from 16 iconic Jewish screen characters: how often their language turns to money, charity, faith, and force.

Film Television Jewish representation Stereotypes Dialogue analysis

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Most of us carry a picture of the Jewish character on screen, assembled from decades of watching, and that picture has a shape: shrewd with money, neurotic, defined by community. The obvious question is whether the actual words those characters speak match the image. We read 47,534 words of speaker-attributed dialogue across 16 iconic Jewish leads and measured them on ten categories of language, and the dialogue refuses any single profile: the sharpest divider in the corpus is not money talk but first-person reference, how often a character says "I," "me," and "my," which separates the cast more (a standardized spread of 1.58) than the money-versus-charity contrast that carries the best-known stereotype (1.27).

This report measures the language of the writing. It makes no claim about any character's morality, about the people who wrote these films, or about Jewish people. The corpus is a canonical, illustrative set, not a representative sample of every Jewish character on screen.

Key Findings

- Measured across ten word categories, the dialogue disperses on every axis rather than settling into one profile. The sharpest single divider is not a stereotype but the share of first-person reference: I, me, and my separate the cast more (a standardized spread of 1.58) than the money-versus-charity contrast does (1.27). (Hanover Institute analysis, 2026.)
- Money-related words are densest in the dialogue of Shylock (17.3 per 1,000 words) and Itzhak Stern (16.6), the moneylender of a Shakespeare play and the bookkeeper of Schindler's List, which shows the count marks a script's subject, not greed. (Hanover Institute analysis, 2026.)
- The single most recognized Jewish screen lead sits near the bottom on money. Alvy Singer of Annie Hall ranks last of the 16 on money words (2.4 per 1,000) and first on intellect words (30.9). (Hanover Institute analysis, 2026.)
- Religion-language concentrates where a character's Jewish identity is the story's crisis. Danny Balint of The Believer, written as a Jew who becomes an antisemite, carries both the highest religion-word density (18.1 per 1,000) and the highest force-word density (11.7). (Hanover Institute analysis, 2026.)

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

- Each character's most over-represented words name its particular world, not a shared trope: a jeweler's gem and auction, a survivor's certificate and captain, a shtetl father's sabbath and tradition. (Hanover Institute analysis, 2026.)

Do these characters share a verbal profile?

No. Placed by how much the dialogue invokes money against charity, the 16 scripts scatter across the whole plane rather than gathering in one corner. The giving end holds Tevye of Fiddler on the Roof (24.8 charity words per 1,000) and the survivor Wladyslaw Szpilman of The Pianist (20.0); the money end holds Howard Ratner of Uncut Gems (15.5); and most of the cast sits in the middle. No region of the chart is the Jewish region.

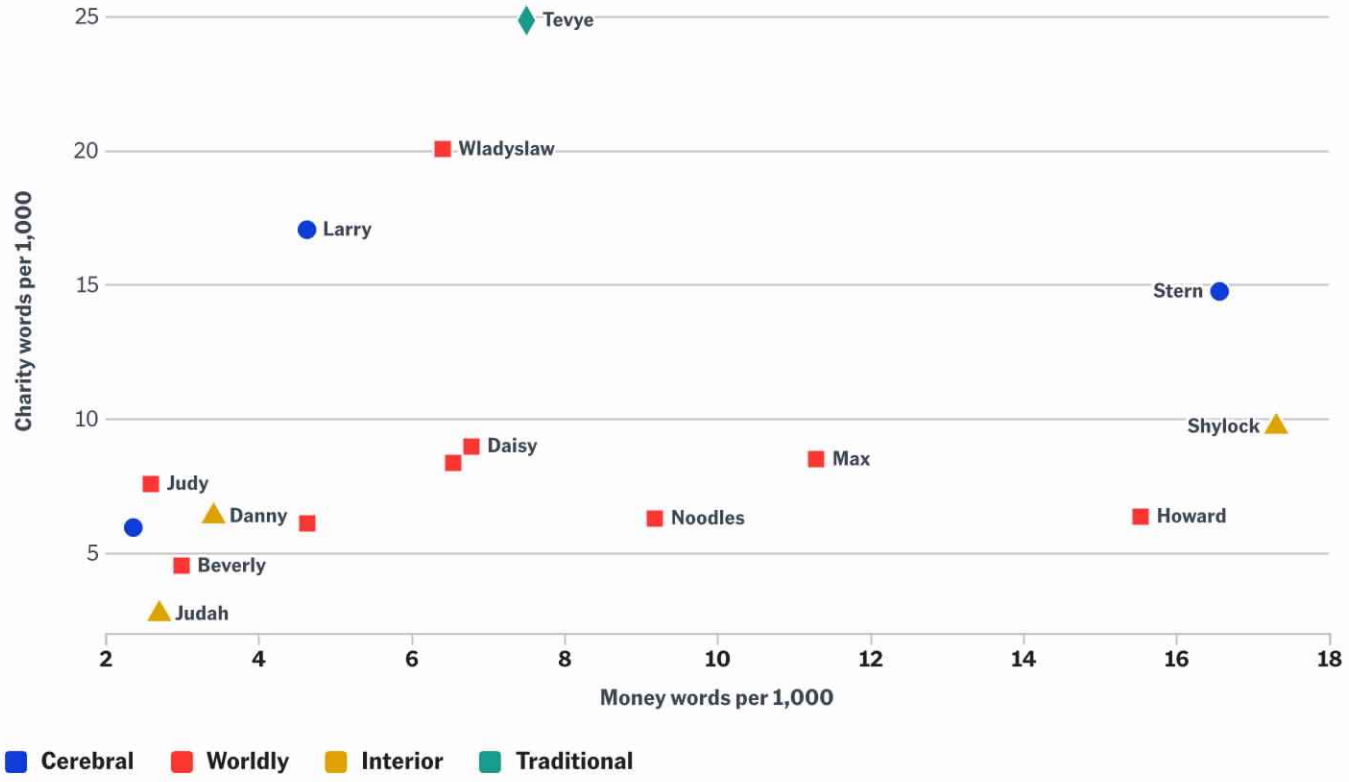
The placement that matters most is Itzhak Stern's. He sits at the top right, high on money words and high on charity words at once, the only character extreme on both. The accountant talks constantly about money because the plot is an act of bookkeeping, and constantly about giving because the bookkeeping saves lives. A character can be saturated with the vocabulary of the trope and sit nowhere near its meaning.

Money language against charity language, by character

THE LEADS SCATTER RATHER THAN CLUSTER ON ANY SINGLE PROFILE

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM



Source: Hanover Institute analysis of 47,534 words of screen dialogue across 16 leads, 2026. Densities are matched lexicon words per 1,000 spoken words.

How far apart do money words sort two characters at the same density?

Not at all. Money-related words are densest in the dialogue of Shylock (17.3 per 1,000) and Itzhak Stern (16.6), whose scripts could hardly be more different: one is the moneylender at the center of a revenge plot, the other an accountant whose bookkeeping protects a factory's Jewish workers. Howard Ratner follows at 15.5, then Max Bialystock of The Producers at 11.3. The density counts how often a script's subject is money; it does not measure greed, and it does not sort these characters by it.

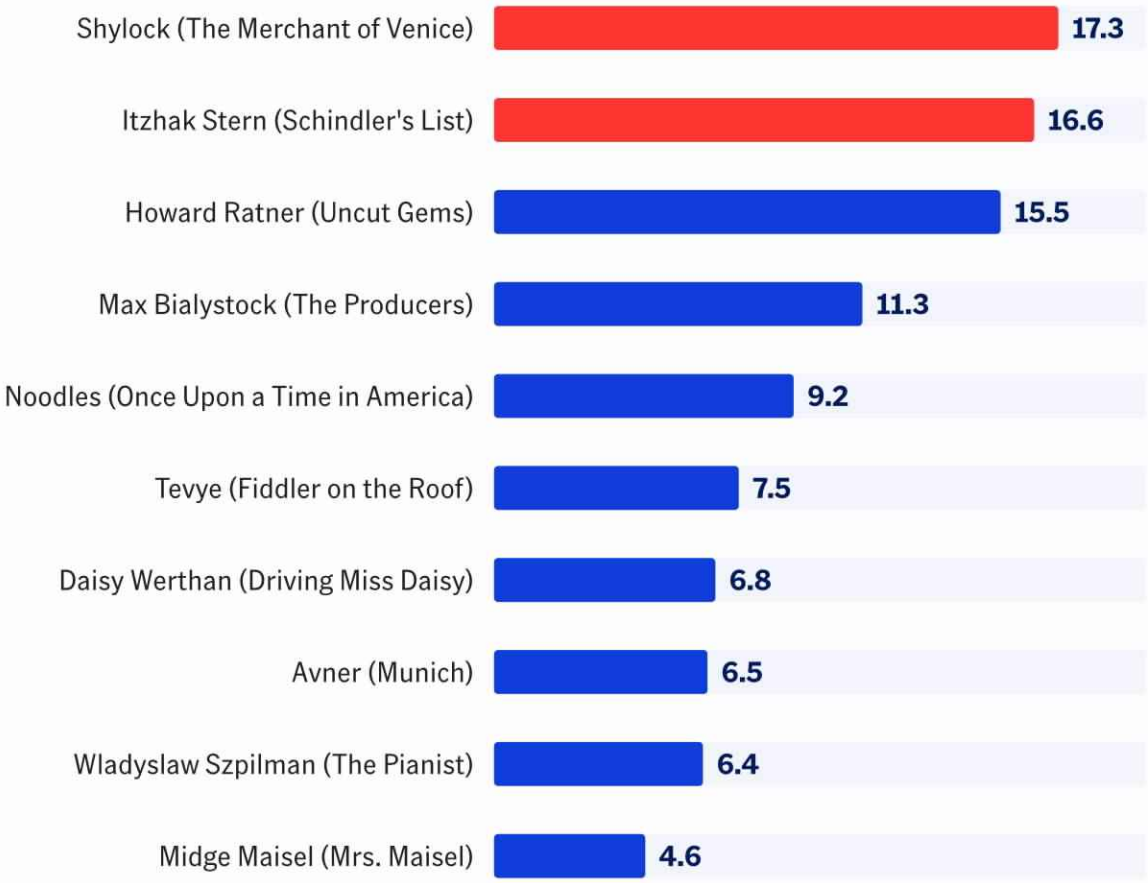
Below the top of the list, money words thin out fast. Alvy Singer, Judy Benjamin, and Judah Rosenthal all sit near 2.5 per 1,000, roughly one-seventh of Shylock's rate. The trope-laden vocabulary is present in a few of these characters and nearly absent in most, which is the opposite of a shared signature. The same caution applies in reverse to small samples: Stern's dialogue is the briefest in the corpus at 543 words, so his rate rests on a thin base and is read here only against the broad pattern, not as a precise point.

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

The same money-word density at the top, opposite jobs in the script

SHYLOCK AND STERN LEAD THE COUNT; IT MARKS A SCRIPT'S SUBJECT, NOT GREED



Source: Hanover Institute analysis of 47,534 words of screen dialogue across 16 leads, 2026. Densities are matched lexicon words per 1,000 spoken words.

What divides the cast more than any stereotype?

A grammatical habit, not a thematic one. Of six opposed word-pairs measured, the contrast that spreads the cast the most is self-reference against other-reference, how often a character says I, me, and my versus how often the dialogue turns outward (a standardized spread of 1.58). The confessional and self-narrating leads sit at one end (Judah Rosenthal, Alvy Singer, Shylock); the characters whose talk is organized around other people sit at the other (Stern, Tevye, Szpilman). The money-versus-charity contrast that carries the famous stereotype ranks only third, at 1.27, behind the sacred-versus-appetite contrast at 1.53. The weakest divider of all is family against ambition, at 0.81: on that axis the cast barely separates.

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

The order is the finding. If a single stereotype governed how these characters are written, a thematic axis like money would dominate the separation. Instead the strongest divisions are a matter of voice and preoccupation, the same dimensions that distinguish any well-drawn character from another.

Which contrasts divide the cast, and which do not

SPREAD OF EACH OPPOSED-WORD-PAIR ACROSS THE 16 LEADS (HIGHER = MORE SEPARATING)

Source: Hanover Institute analysis of 47,534 words of screen dialogue across 16 leads, 2026. Densities are matched lexicon words per 1,000 spoken words.

Where does religious language sit against the language of the body?

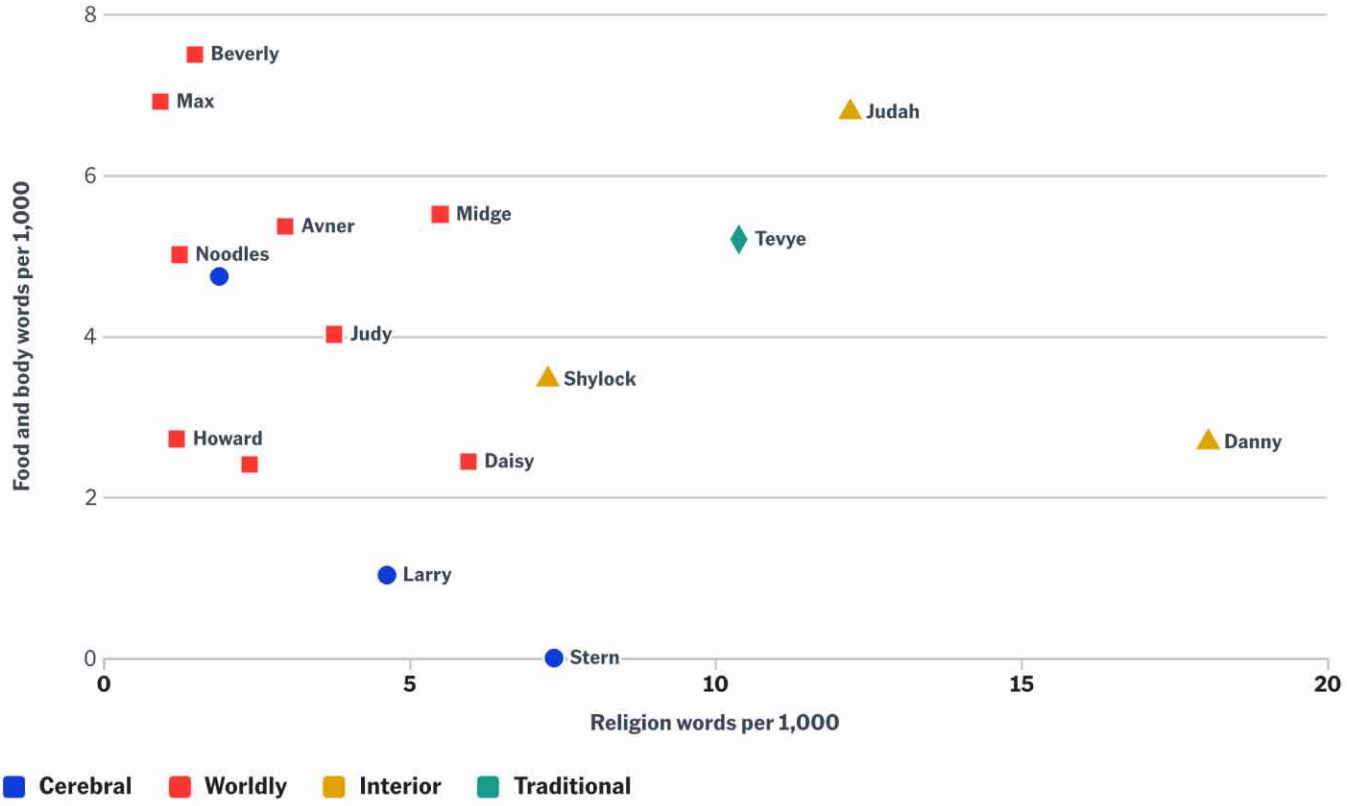
Religion-words and food-and-body words pull apart cleanly, and the split tracks a character's relationship to its own tradition rather than any group trait. Danny Balint, written as a yeshiva-raised Jew who turns against his own identity, carries the highest religion-word density in the corpus (18.1 per 1,000); Judah Rosenthal (12.2) and Tevye (10.4) follow. At the other pole sit the worldly and domestic characters whose dialogue runs to food and the body: Beverly Goldberg, Max Bialystock, and Noodles. The sacred register concentrates where the plot is about faith and identity, and recedes where it is about getting through the day.

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Religion-words against food-and-body words, by character

THE SPLIT TRACKS A CHARACTER'S RELATIONSHIP TO ITS TRADITION



Source: Hanover Institute analysis of 47,534 words of screen dialogue across 16 leads, 2026. Densities are matched lexicon words per 1,000 spoken words.

Who fights and who reasons?

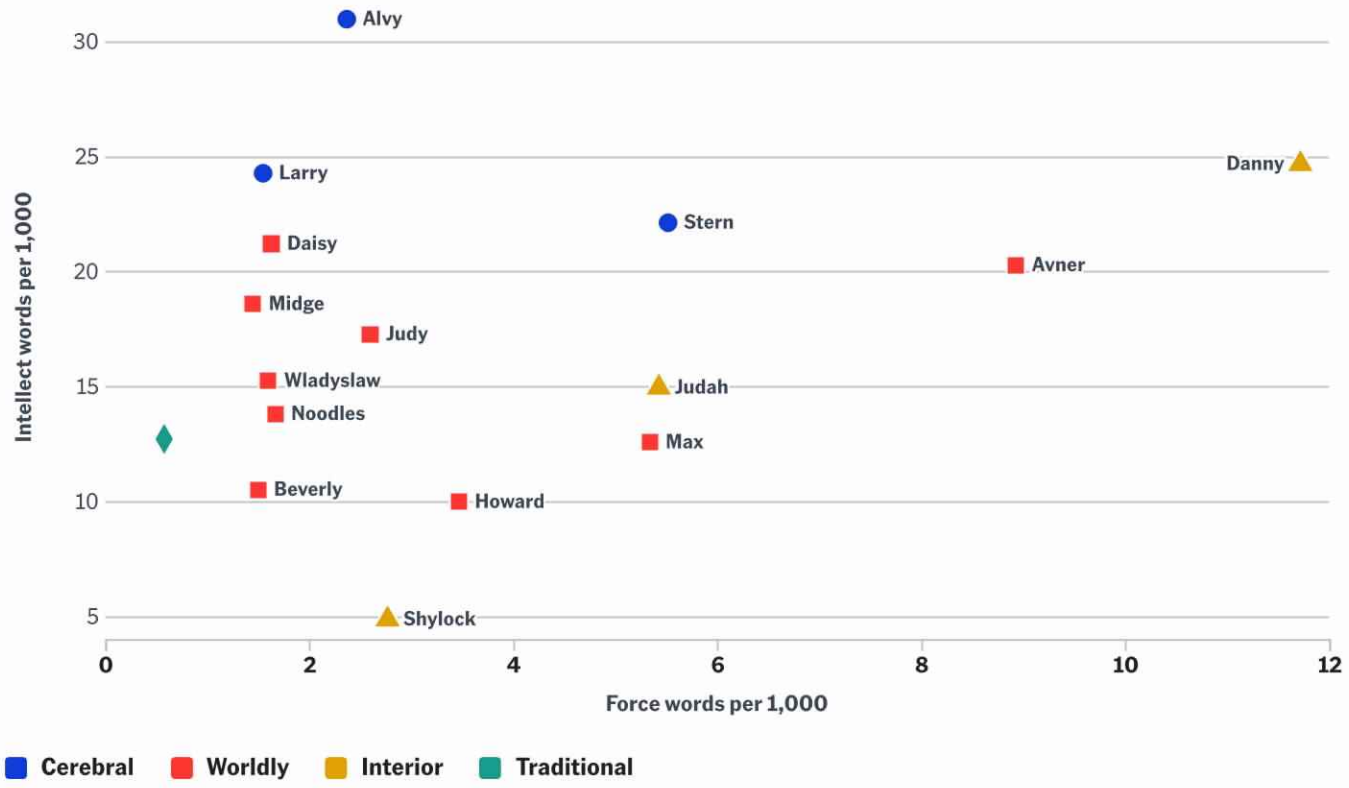
The violence-and-mind contrast separates the action characters from the talkers, and again the placement cuts across any single profile. Alvy Singer anchors the intellect corner (30.9 intellect words per 1,000, 2.4 force words), with Larry Gopnik and Daisy Werthan near him. The force end holds the gangster and the operative, Noodles and Avner of Munich, and the verse-bound Shylock. The most telling position is Danny Balint's, alone in the upper right: his dialogue is at once among the most force-laden and among the most intellect-laden, the two registers running high together.

Force-words against intellect-words, by character

WHERE THE ACTION CHARACTERS AND THE TALKERS FALL

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM



Source: Hanover Institute analysis of 47,534 words of screen dialogue across 16 leads, 2026. Densities are matched lexicon words per 1,000 spoken words.

What words set each character apart?

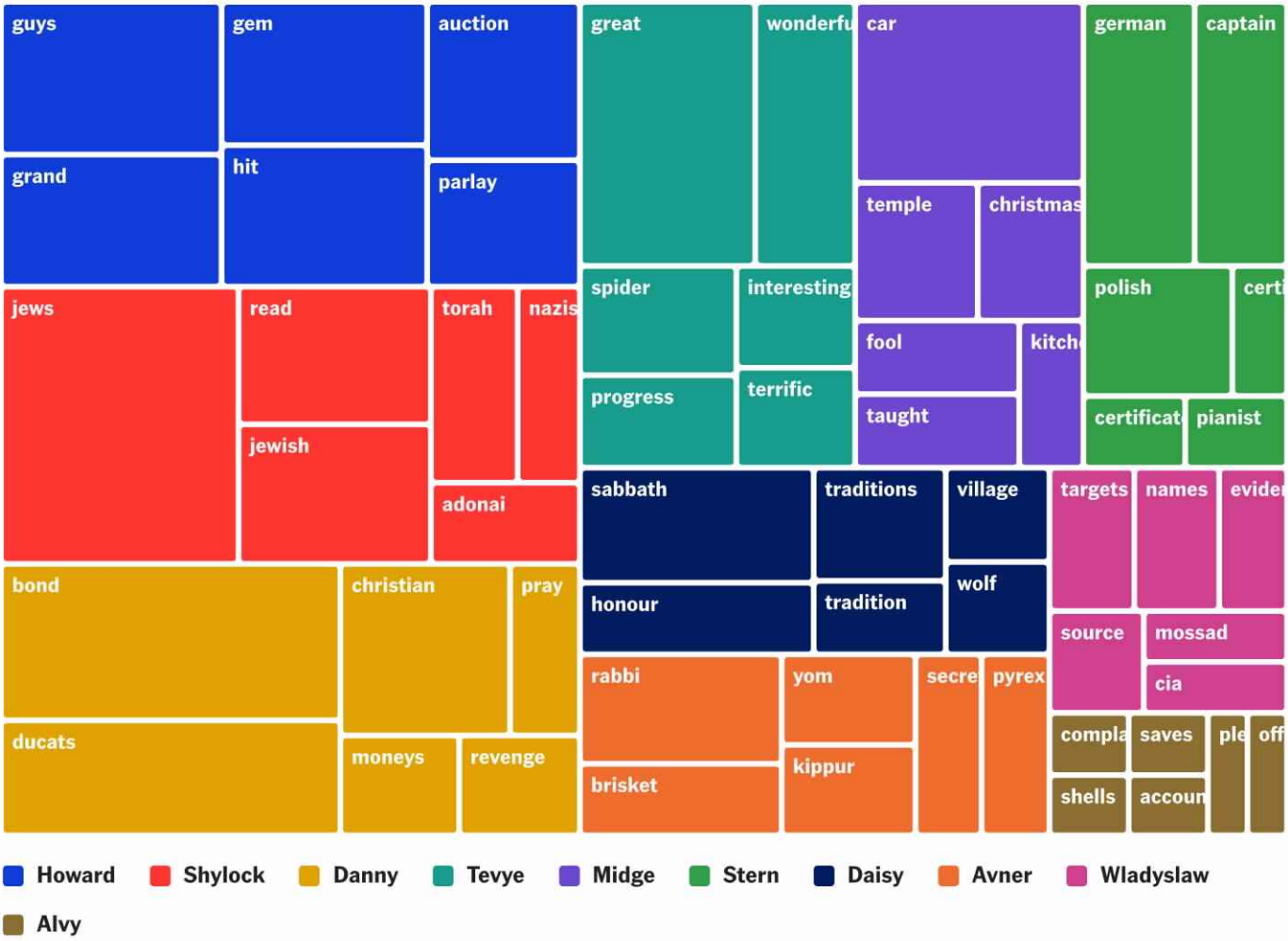
The distinctive vocabulary of each lead, the words it uses far more than the others do, names the particular world of its film rather than a common stereotype. Measured by term frequency against the rest of the corpus and stripped of proper names, Howard Ratner's signature words are grand, gem, auction, and parlay; the survivor Szpilman's are certificate, captain, German, and Polish; the shtetl father Tevye's are sabbath, tradition, and honour; Midge Maisel's are brisket, yom, kippur, and pyrex. Danny Balint is the exception that proves the point: his most distinctive words, jews, torah, nazis, and jewish, are the explicit vocabulary of an identity at war with itself, the one character for whom Jewishness is the literal subject of nearly every line.

The words that set each lead apart

DISTINCTIVE VOCABULARY BY TF-IDF, PROPER NAMES REMOVED; EACH BLOCK IS ONE CHARACTER'S OWN FILM DIALOGUE

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM



Source: Hanover Institute analysis of 47,534 words of screen dialogue across 16 leads, 2026. Densities are matched lexicon words per 1,000 spoken words.

Methodology and limitations

This report presents an original Hanover Institute study. The unit of measurement is the word, and every figure is an aggregate over a character's complete spoken dialogue; the study measures the language of the writing, not any character's morality, the people who wrote these works, or Jewish people.

Provenance. The corpus is 16 lead Jewish characters from American film and television, selected as widely recognized, canonical figures whose Jewishness is established in the work and whose screenplays carry speaker-attributed dialogue. Dialogue was isolated per character from publicly available scripts. For works in the public domain (the Shakespeare text), the verbatim dialogue was processed directly; for in-copyright works, only an

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

aggregate word-frequency table was computed and retained, and no verbatim dialogue was stored, so the study holds statistics about the texts rather than the texts themselves.

Sample. The corpus totals 47,534 spoken words. Per-character samples range widely, from 8,434 words (Alvy Singer) down to 543 (Itzhak Stern), 667 (Beverly Goldberg), and 737 (Judah Rosenthal). Densities for the smallest samples are noisier and are read here against the overall pattern, not as precise points. One character, Shylock, speaks in verse, a different register from screen prose.

Enrichment. Each character's dialogue was scored on ten transparent lexical categories (money, charity, family, ambition, anxiety, intellect, religion, body, force, and self-reference), reported as matched words per 1,000 spoken words. The word lists are fixed and documented in the study files, so any reader can check what each category counts. Distinctive vocabulary was ranked by term frequency-inverse document frequency against the rest of the corpus, with proper names removed. The four color groups on the scatterplots come from a k-means clustering on the standardized category scores, and the opposed-pair spreads are standardized differences, so neither pole of a contrast is favored by its baseline rate.

Limitations. The corpus is a canonical, illustrative selection, not a representative sample of all Jewish characters on screen, and shares are not generalizable beyond it. The study has no non-Jewish comparison group; a paired Jewish and non-Jewish corpus is the natural next study. Word frequency is a measure of a script's subject matter, not of a character's personality or any group's traits: a high money-word count marks how often money is discussed, not greed. The lexical categories are author-chosen and some overlap by design (money and ambition share vocabulary). Optical-character-recognition and dialect transcription introduce minor noise, and several scripts were parsed from formatted PDFs.

Prior research is thin and almost entirely interpretive. The one prior quantitative count of Jewish characters is the USC Norman Lear Center's Jews on Screen (2024), a content analysis of 108 Jewish characters across 15 recent US scripted television series. It coded four stereotype clusters that overlap this study's categories, the Greedy Shylock, the Wealthy and Powerful, the Nebbish, and overbearing parents, and found that just 18% of those characters (19 of 108) referenced Judaism in their dialogue, a pattern consistent with the low religion-word densities measured here for most of the cast. The critical canon, including Patricia

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Erens's The Jew in American Cinema (1984), Lester D. Friedman's Hollywood's Image of the Jew (1982), and Henry Bial's Acting Jewish (2005), names the archetypes this study measures through close reading, but does not measure dialogue. This report adds word-level measurement across a fixed corpus, a quantitative layer absent from that critical literature.

Conclusion

So does the language match the memory? The 16 scripts refuse the single stereotype the image promises, and yet a portrait does rise out of the numbers, just not the expected one. These characters are written as minds, not bodies: none is an action hero, the cast runs far richer in intellect than in force, and its sharpest internal divide is not money but how often a character turns inward to say I, me, and my. They are as often giving as grasping; the charity end of the money axis holds Tevye and the survivor Szpilman. Where money-words do pile up, they pile up on a moneylender and a life-saving bookkeeper alike, marking a script's subject, not a soul's greed.

So the stereotype's vocabulary is everywhere in these films and the stereotype itself is nowhere, and that gap is the finding. Yet a shape remains, the one most viewers carried in: brilliant and verbal, often generous, just as often neurotic and uneasy about money, and, on the evidence of the dialogue, almost never the hero who simply prevails. The data holds the virtues plainly, these are among the most intellect-rich and most charitable leads one could gather, which sharpens a question rather than settling it: why is that talent so rarely written as a leading man who triumphs on it? Flaws make a character human; the open question is why these particular flaws are the ones the culture reaches for, on these particular characters.

That question does not stay in the theater. The image of the money-minded Jew lives well outside it, in everyday jokes, in conspiracy theories about who controls what, in asides dropped for a laugh, a stereotype normalized far beyond any film. Cultural portrayal is one of the inputs that shapes how a group is seen. Whether a long habit of writing Jewish leads as clever and money-adjacent rather than heroic is one of the quiet feeders of that stereotype is not a question a word count can answer. What the count can show is the portrayal itself: minds written richer than bodies, generosity running as deep as anxiety, and heroism nearly absent from the ledger of eight decades of film.

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Frequently Asked Questions

How were the 16 characters chosen?

They were selected as widely recognized, canonical Jewish leads from American film and television whose Jewishness is established in the work and whose screenplays carry speaker-attributed dialogue. The selection is deliberate and illustrative, not a random or representative sample, which is why the report draws no population-level shares from it.

Does a high money-word count mean a character is greedy?

No. The count measures how often a script's subject is money, not a character's motive or morality. The clearest evidence is that the two highest money-word densities belong to Shylock and to Itzhak Stern, whose bookkeeping in Schindler's List is used to protect Jewish workers; the same vocabulary serves opposite ends.

Why include Shylock, who is not a film character?

Shylock is the source of the moneylender archetype the later characters are measured against, so the corpus would be incomplete without him. His dialogue is verse rather than screen prose, a different register that the methodology notes; the public-domain text also allowed direct processing.

Which kinds of language divide the cast the most?

Self-reference against other-reference divides it most (a standardized spread of 1.58), followed by sacred-versus-appetite language (1.53). The money-versus-charity contrast that carries the best-known stereotype ranks third (1.27), and family-versus-ambition barely separates the cast at all (0.81).

Does this study measure antisemitism in film?

No. It measures the language of a set of canonical Jewish characters. The Institute treats screen representation as one input that can shape perceptions of a group, and studies it descriptively; this report documents how these characters are written, not whether any portrayal is prejudiced.

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Could the same method compare Jewish and non-Jewish characters?

Yes, and that is the natural next study. Run on a matched corpus of non-Jewish leads, the ten-category method would separate the patterns specific to how Jewish characters are written from the patterns common to screen dialogue in general.

Sources

- Bial, Henry, 2005. Acting Jewish: Negotiating Ethnicity on the American Stage and Screen. Ann Arbor: University of Michigan Press. ISBN 9780472069088. Scholarly book.
- Erens, Patricia, 1984. The Jew in American Cinema. Bloomington: Indiana University Press. ISBN 9780253145000. Scholarly book.
- Friedman, Lester D., 1982. Hollywood's Image of the Jew. New York: Frederick Ungar Publishing. Scholarly book.
- Hanover Institute, 2026. The Language of Iconic Jewish Screen Characters: A Lexical Analysis of 16 Leads. Original study; 47,534 words of speaker-attributed dialogue across 16 characters. Institute research.
- USC Norman Lear Center, 2024. Jews on Screen: The Visibility and Representation of Jewish Identity in Contemporary Scripted TV. Content analysis of 108 Jewish characters across 15 US scripted television series. learcenter.org. Established research organization.

RELATED RESEARCH

DISCOURSE JUNE 25, 2026

How Gaza-War Headlines Name the Agent of Violence

A corpus study of 1,976 Gaza-war headlines: only 19% depict a violent act, and when one names an actor it is Israel 80% of the time and no one 23%.

EXPLAINER JUNE 26, 2026

What Is Antisemitism? Reading the Language It Hides In

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Antisemitism is mostly written in code, not slurs. This guide decodes the coded lexicon, shows how it evades detection, and tracks its measured rise.

HANOVER INSTITUTE

The Hanover Institute for Public Policy studies the inputs fueling antisemitism in the United States and publishes what the evidence shows. Its work is content analysis, computational social science, and other data-driven research, and its reports take no policy positions.

SUBSCRIBE TO OUR NEWSLETTER

Your email address SUBSCRIBE

RESEARCH

Research

Methodology

INSTITUTE

About

Mission & Standards

Contact

TOPICS

Online & Social Media

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM

Narratives & Tropes

Media & Representation

Israel & Anti-Zionism

AI & Technology

Incidents & Trends

Drivers & Effects

FORMATS

Data Report

Explainer

Narrative Analysis

Method

Trend Analysis

Comparison

Discourse

This material is distributed by Piro, Inc. on behalf of Havas Media Germany GmbH on behalf of the Israel Government Advertising Agency (LaPam). Additional information is available at the Department of Justice, Washington, DC.

© 2026 The Hanover Institute for Public Policy

Terms of Use Privacy Policy

Received by NSD/FARA Registration Unit 08/07/2026 6:13:15 PM