

Does System Matter (in TTRPG Conversation)?

Game master, system and players at sixty recorded tabletop role-playing tables

Valentin Thouzeau^{1,✉} and Matthieu Minne²

¹ Université Paris Sciences et Lettres, Paris, France · ² Independent researcher, Paris, France

✉ Corresponding author: valentin.thouzeau@gmail.com

Working paper, 19 September 2026 (v9, English version)

Abstract. A tabletop role-playing game (TTRPG) session is a conversation: a game master (GM) describes a world and players make their characters act in it. What gives this conversation its shape, and why do two tables not play the same way? We study 60 French-speaking tables that published their sessions as video (200 sessions, 337 hours, 12 systems, 48 GMs). An automatic chain transcribes the speech, recognises who is speaking and codes every turn as speech acts, by what the speaker does and what the act bears on (the fiction, a character's words, the rules, the life of the table). Three regularities do not depend on the game: the GM holds more than half of the speaking time; what the GM lays down in the fiction and what the players answer form the centre to which everything returns; and the conversation alternates between four loops, one per level. By comparing the act-to-act sequences of every pair of tables, corrected for the number of acts observed at each table, we apportion the differences between tables. The GM accounts for a third (34%), the system for a quarter (23%), the number of players for 6%, the share of women for nothing measurable (2%), and a third remains specific to each table. The order between GM and system is not settled: the same GMs rarely run several systems, and the GM cannot be told apart from the channel that publishes them.

Keywords: role-playing games, speech acts, conversation analysis, corpus, automatic annotation, hidden Markov model, variance components, multiple regression on distance matrices.

1. Introduction

1.1. Role-playing as conversation

Tabletop role-playing is an activity made almost entirely of speech. A game master describes a world and brings its characters to life; players each make their own character act in it; rules, dice and character sheets come into play at the moments that the system of the game defines. What is said there belongs to several frames at once, in Goffman's (1974) sense: participants shift constantly from person to player and from player to character (Fine, 1983), and the frame itself is spoken, as in children's make-believe (Giffin, 1984; Sawyer, 1997). The fiction that results is a narration with several voices, whose course is organised by the participation of those who listen to it (Goodwin, 1984).

This conversation has a property that ordinary conversation lacks: one participant holds an instituted role. The GM narrates, arbitrates, voices the characters that belong to no one, and distributes information and the floor (Tychsen, Hitchens, Brolund, & Kavakli, 2005; Hendricks, 2006; Cover, 2010). And a session is played under a system: a game, with its rules, and often a published scenario, which fixes what is talked about.

1.2. Describing a conversation by its acts

Since Austin (1962) and Searle (1969, 1976), an utterance is described not only by what it says but by what it does: asserting, asking, promising, declaring. This description suits role-playing particularly well, since most utterances there make sense only through what they do to a shared state: "My character walks up to the goblin" would have no effect outside the game, and inside it means that a new fact exists in the common fiction. Linguistics has formalised this description in taxonomies of dialogue acts (Core & Allen, 1997; Bunt et al., 2012), after Bales's (1950) scheme of twelve categories of acts in groups, whose proportions and sequences revealed phases of the discussion (Bales & Strodtbeck, 1951).

Role-playing is an institution of its own. The participants are not equal: the GM has authority over the world and the rules, each player over their character. The objects of speech are proper to it as well: a fiction that exists only through what is said of it, characters whose words are uttered live by the people who play them, rules and dice handled while speaking, and the ordinary life of the people around the table. We therefore describe each act on two dimensions that the institution itself establishes. Its **force** is what the speaker does in speaking: asking, for instance, or attempting a

move whose outcome the GM or the dice will decide. Its **level** is what the act bears on: the fiction, a character's words, the game apparatus, or the life of the participants outside the game. The five forces and the four levels are defined in Section 4.1; they adapt the classification of Searle and Vanderveken (1985) to what the game requires one to distinguish, and take up Fine's (1983) frames and Clark's (1996) layers of real and represented action.

Described in this way, a session is a sequence of acts, and practitioners' descriptions and ethnographies agree that this sequence is not arbitrary: certain patterns recur incessantly, a player's question followed by the GM's answer, a character's action followed by its consequences, a dice roll followed by its result and its translation into the world, the lines that characters exchange. We call a **loop** a phase of the conversation during which acts mix in the same way, which recurs several times over a session and which transitions separate from the others. Example: in a work meeting, a report by one participant, a round of questions and answers, a debate and a digression about the coffee machine are four such phases, and everyone notices when the conversation moves from one to another. The working hypothesis is that a session is composed of a small number of loops and that most acts sit within them. Two results make it plausible: the sequence of acts in a conversation is predictable from the few acts that precede it (Stolcke et al., 2000), and a group discussion passes through phases that can be read in the proportions of acts (Bales & Strodtbeck, 1951).

1.3. What might structure the conversation

If a session is made of loops, what governs their mixture, their duration and their sequence, and why do two tables not play the same way? Four candidates are open to measurement in a corpus of published sessions.

The first is the **system**, that is, the game and the scenario played. That system matters is the founding statement of game designers' theory (Edwards, 1999; Torner, 2018); Edwards's essay is aimed at the idea that any system works with a good GM, and the opposition between GM and system that our title takes up is his. A system there is the way the group distributes the right to assert what (Young, 2005) and organises what each participant may or may not say (Sintes, 2013). For Briand (2014), the game system is the set of means by which the real participants agree to move the imagined situations forward, who describes what, when and why; the arrangement in which one GM faces the players is one such system among others, and sharing narrative authority is a design choice with decisive consequences. For Baker (2010, 2019), who designed the family of games "powered by the apocalypse", play-

ing is a conversation and the rules are what arbitrates it, through moves triggered by what the character does in the fiction. If these views are right, the system should leave a measurable imprint on the sequences of acts, in particular on the frequency and duration of passages through the rules.

The second is the **GM**. Ethnographies and discourse analyses describe the GM as the one who holds the frame, narrates, ratifies and distributes the floor (Fine, 1983; Hendricks, 2006; Cover, 2010), and practitioners as the one whose decisions give a table its style (Kim, 1998) and whose skills carry over from one game to the next (David & Larré, 2016). In multiparty conversation, selecting the next speaker is an operation in itself (Lerner, 2003); the GM performs it constantly (“What do you do?”), in a position comparable to that of the teacher who initiates, obtains an answer and evaluates it (Sinclair & Coulthard, 1975; Mehan, 1979). The GM’s way of doing things might stamp the conversation with a signature independent of the game played.

The third is the **number of players**. The mechanics of turn-taking (Sacks, Schegloff, & Jefferson, 1974) suggest an arithmetical effect: the more players there are, the more a player’s act is followed by another player’s act rather than by the GM’s. The sociology of small groups has shown that the first speaker dominates all the more as the group is larger (Bales, Strodtbeck, Mills, & Roseborough, 1951; Stephan & Mishler, 1952). Who speaks after whom follows rules specific to the group (Parker, 1988; Gibson, 2003), and the form of a discussion changes with its size (Fay, Garrod, & Carletta, 2000).

The fourth is the **gender** composition of the table, which voice identification makes measurable. Role-playing is a gendered hobby: the groups that Fine (1983) observed were almost exclusively male, and the gaming table has been described as a male preserve whose talk is coloured by that dominance (Dashiell, 2020). In general, however, gender differences in the amount and type of talk are small and context-dependent (James & Drakich, 1993; Leaper & Ayres, 2007; Karpowicz & Mendelberg, 2014), and nothing of the kind has been measured at a role-playing table.

We study the form of the conversation, the proportions of acts and transitions, and not its content: form is what regulates the social structure of a group independently of what is said (Koudenbourg, Postmes, & Gordijn, 2017), and it is what an automatic chain measures most reliably.

1.4. A corpus of published sessions and an automatic chain

Empirical work on role-playing conversation has so far rested on small corpora read by hand: a few tables observed for months (Fine, 1983), a few sessions (Hendricks, 2006; Cover, 2010; Tychsen, Smith, Hitchens, & Tosca, 2006), an hour and a half of one episode transcribed according to the conventions of conversation analysis (Sharp, 2023). Two things have changed. Recorded and published sessions, which the community calls *actual play*, have become a media genre in their own right (Roux, 2020; Chalk, 2023; Jones, 2024), and hundreds of hours of play are available for every popular scenario. Large corpora of sessions exist, built on transcriptions made by fans or on play by writing (Rameshkumar & Bailey, 2020; Callison-Burch et al., 2022; Zhu, Aggarwal, Feng, Martin, & Callison-Burch, 2023), and automatic transcription extends this possibility to spoken recordings. And large language models annotate texts about as well as human annotators recruited online, at very low cost (Gilardi, Alizadeh, & Kubli, 2023; Ziems et al., 2024), including dialogue acts coded from a codebook (Ostyakova, Smilga, Petukhova, Molchanova, & Kornev, 2023). Game dialogue corpora exist for board games (Asher, Hunter, Morey, Benamara, & Afantenos, 2016; Boritchev & Amblard, 2022), but not for spoken role-playing.

This article describes a French-language corpus built for the question at hand: several tables playing the same published scenario, several scenarios, a few GMs present at several tables, so that the

GM, the system and the number of participants vary as separately from one another as the published material allows. The article describes the chain that leads from the published video to that apportioning, in the order of the chain (Figure 1): each section describes a step, then what it reveals. The main text is written to be read without any statistical background; the technical details are in the appendices.

The chain yields three kinds of result, of unequal status. On the way, Sections 3 to 6 describe the corpus: the GM’s share of speech, the graph of the transitions from one act to the next, and the loops; these descriptions hold whatever the system. At the end of the chain, Sections 7 and 8 answer the question: the GM accounts for about a third of the differences between tables, the system for a quarter, the number of players for a small part, the share of women for nothing measurable, and a third belongs to each table. Section 9 then asks, transition by transition, where each factor acts, and follows one case down to the words exchanged; that section is exploratory. Section 10 discusses.

2. The corpus

2.1. Collection

The corpus is made of sessions published on YouTube by French-speaking channels, chosen because they play the same commercial scenarios: *The Haunting* (*Call of Cthulhu*; played in French as *La Maison Corbitt*), *Chariot of the Gods* (*Alien*; *Le Chariot des dieux*), *Lost Mine of Phandelver* and *Death House* (*Dungeons & Dragons*, fifth edition), the latter played once in *Daggerheart*. To compare these classic games with another family of rules, campaigns of games “powered by the apocalypse” or derived from them, played by the same channels, were added (*Monster of the Week*, *Blades in the Dark*, *Candela Obscura*, *Root*, *The Sprawl*, *Avatar Legends*, *Masks*); we call them below the “apocalypse” games. The collection was done from the platform’s public pages, without an account; only the audio is kept. Games and scenarios are referenced in the ludography, after the references.

2.2. Vocabulary and units

A **system** is a game and a scenario: a *Dungeons & Dragons* table playing *Lost Mine of Phandelver* and another playing *Death House* play two systems, and *Death House* played in *Daggerheart* is a third. For the classic games, the systems are thus the four published scenarios, plus *Death House* in *Daggerheart*; for the “apocalypse” games, the game alone, each campaign having its own scenario. A **table** is a group playing a system; it is the unit of the whole analysis, and all its sessions are pooled. The **GM** is identified by voice and not by the name of the channel, several channels changing GM from one campaign to the next. The **number of players** is the number of player voices present in at least 2% of the turns. The **share of women** is the share of participants (GM included) whose voice is classified as female (Section 3.3).

2.3. What the corpus contains

Table 1 describes the corpus: 60 tables, 200 sessions, 337 hours, published between 2014 and 2026 by 34 channels, 12 systems, 48 GMs. A table has from 1 to 14 sessions (median 3), from 1.3 to 13.6 hours (median 5.1) and from 418 to 5,407 annotated acts (median 1,832). The number of players ranges from 2 to 6 (mean 4.35). Women are 24% of the 325 enrolled voices; 18 tables have none, and two GMs out of 48 are women. System and GM are nested: 41 GMs out of 48 have a single table. The other 7 are the GMs of channels that publish several campaigns (the 48 GMs and their tables are listed in Table A1). Their 19 tables form the 18 pairs of tables with the same GM on which everything the article says about the GM rests. A sixty-first table was discarded before the analysis because the roles of GM and players could not be reliably attributed there; the descriptions of Sections 4 to 6

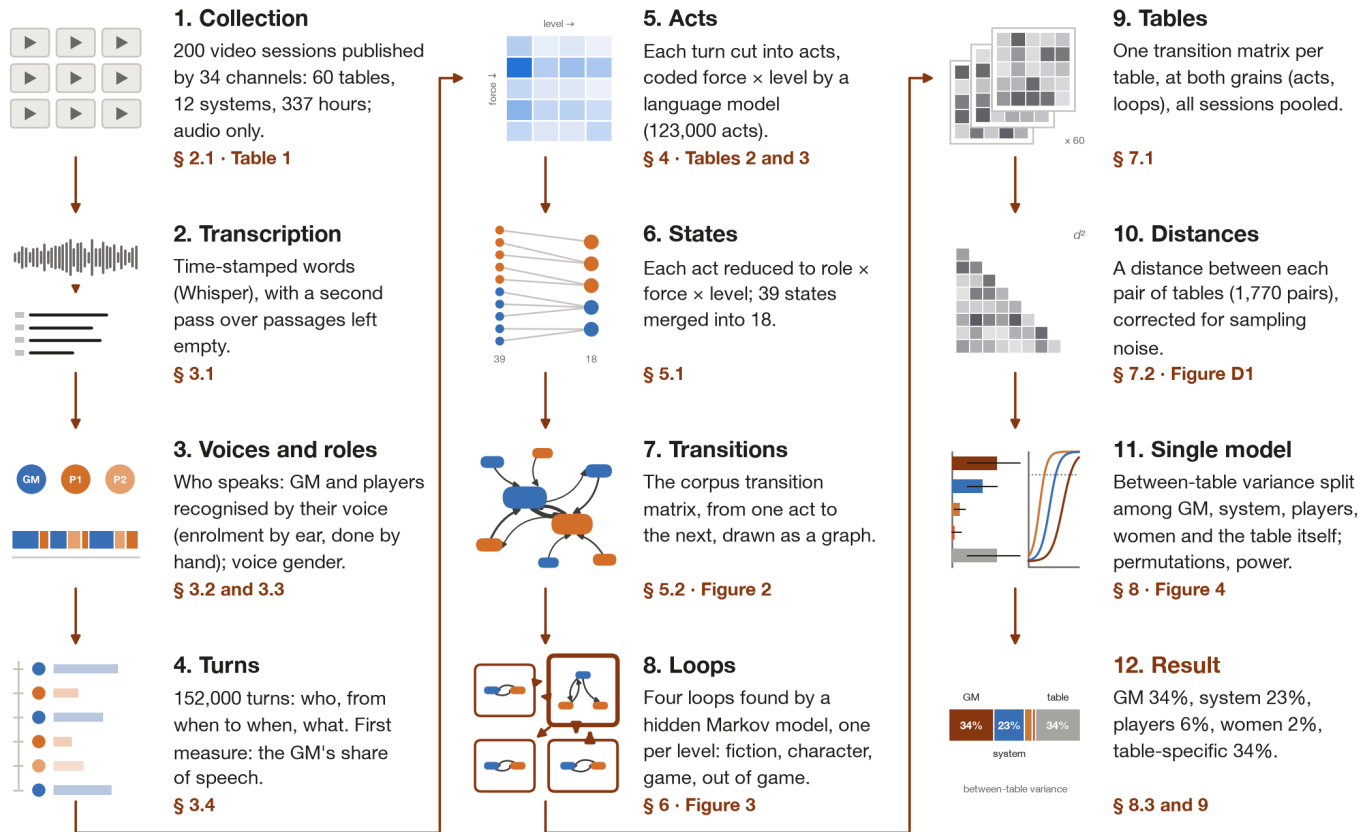


Figure 1. The processing chain, from the published video to the apportioning of between-table variance, in twelve steps read top to bottom, column after column. Each step is represented by a pictogram that takes up, in simplified form, what it produces or the figure of the article that shows it (blue: the GM; orange: the players; brown: the loops and the variance shares), followed by what the step does, the section of the article that describes it and, where relevant, the table or figure that presents its result. The enrolment of voices (step 3) is done by hand.

(Tables 2 and 3, Figures 2 and 3) are computed on the 202 annotated sessions, discarded table included, and everything from Section 7 on, on the 60 tables.

3. From video to speaking turns

3.1. Transcription

The audio of each video was transcribed word by word, with the time of each word, by a speech recognition system (Whisper; Radford et al., 2023). This system sometimes invents sentences after a silence (Koenecke, Choi, Mei, Schellmann, & Sloane, 2024) and sometimes stops before the end of a long recording; a second pass re-transcribed the spoken passages that the first had left empty.

3.2. Voices and roles

The next step is to know who is speaking. A program cuts the audio into speech segments and computes for each segment a

voice embedding, a numerical signature that brings together the segments of one and the same person (Appendix A). This automatic clustering is not enough for a role-playing table: players change their voice to speak in character, and these recordings are harder than the meetings on which such programs are developed (Remme & Tang, 2025). On the first session of each table, the proposed groups of segments were therefore examined with the transcription at hand, the GM recognised as the longest voice that narrates, the other voices named; the embeddings of these samples then served to attribute all the segments of all the sessions of the table. This step is the only manual one. The same people across tables were recognised by the similarity of their embeddings.

3.3. Voice gender

The gender of each voice was estimated automatically by three independent acoustic methods (Appendix A); they agree for 92%

Table 1. The corpus by system. Acts are counted after merging repeated acts (Section 5.1). Distinct GMs are counted by system: the same people return on several systems, hence 48 GMs in all and not the sum of the column. D&D 5: *Dungeons & Dragons*, fifth edition.

system	tables	sessions	hours	acts	distinct GMs	players (mean)	share of women (mean, %)
Chariot of the Gods (Alien)	19	74	128.7	49,536	17	4.8	19
The Haunting (Call of Cthulhu)	9	37	44.4	15,656	8	3.7	32
Lost Mine of Phandelver (D&D 5)	9	22	44.4	16,027	9	4.1	24
Death House (D&D 5)	8	21	41.8	15,862	8	4.9	30
Monster of the Week	3	13	15.9	6,070	3	4.0	29
Blades in the Dark	2	7	12.7	3,500	2	4.0	25
Root	2	9	11.9	3,836	2	3.5	22
Candela Obscura	2	3	11.4	3,269	2	3.5	12
Avatar Legends	2	3	7.7	2,751	2	4.5	0
The Sprawl	2	5	7.0	1,928	2	4.0	0
Death House (Daggerheart)	1	3	6.0	2,753	1	5.0	50
Masks	1	3	5.5	1,809	1	4.0	20
total	60	200	337.3	122,997	48	4.35	24

of the voices, and the gender retained is the mean of the three. No check by ear was made.

3.4. Result: the speaking turns and the GM's share of speech

Transcription and voices were merged into speaking turns (who speaks, from when to when, and what is said): 152,000 turns over the 200 sessions. The GM is the first speaker of 196 sessions out of 200 and speaks, at the median, 3.5 times more than the most talkative player at the table. The GM holds the majority of the speaking time: median 57% per table, quartiles 51 and 63%, from 30 to 75% depending on the table. This share of speech does not vary from one system to another beyond what chance would produce (Appendix A), and it decreases with the number of players (correlation -0.50). The dominance of the first speaker is the rule in small groups (Stephan & Mishler, 1952); its magnitude here, one person who speaks more than all the others put together in a conversation of five or six people, is the mark of an instituted role (Tychsen et al., 2005).

4. From turns to acts

4.1. The codebook

Each speaking turn is cut into acts, and each act is coded on the two dimensions of Section 1.2 (Appendix B). The **force** is what the speaker does in speaking. *Attempts*: the speaker plays a move, makes the character do something or performs an operation of the game, and the effect of that move does not depend on the speaker. *Asserts*: the speaker supplies content, a description, a piece of information, an opinion, a narration. *Validates*: the speaker accepts or refuses what another has proposed. *Asks*: the speaker requests information or an action. *Expresses*: the speaker reacts, without laying down or asking anything. The **level** is what the act bears on. *Fiction*: the world being played, spoken about as a participant. *Character*: words uttered live by a character, the participant lending it a voice; the criterion is that a character in the fiction could hear these words. *Game*: the rules, the dice, the character sheets, the virtual tabletop of remote play. *Out-of-game*: the participants' life outside the game, technical matters, schedules, the people themselves. An act is written force-level: "asks-fiction" is a question about the world being played.

Three codes are not acts: *continuation* (the turn continues the same speaker's previous act), *none* (an acknowledgement, "okay", "hmm") and *unreadable* (unusable transcription). The "okay" and

"hmm" by which a participant signals that they follow are thus left out of the sequence, which separates, as Traum and Hinkelman (1992) do, the acts that move the conversation forward from those that only signal understanding. The codebook further fixes an order for settling mixed cases and a dozen conventions (Appendix B); for instance, a dice result announced on its own is attempts-game, and its translation into the world is an asserts-fiction.

4.2. Automatic annotation

Annotating 152,000 turns by hand was not possible. The annotation was entrusted to a language model (gpt-5.4-mini; OpenAI, 2026), which receives the codebook, a hand-annotated passage of 54 turns as an example, and the turns in slices of one hundred, and returns for each turn one or more coded acts. Four models were compared on the same 11 sessions, against an annotation by a more powerful model (Claude Opus 5; Anthropic, 2026) and against two hand-annotated passages; gpt-5.4-mini was retained for its ratio of agreement to cost. As in other tasks entrusted to these models, the dimension that is simplest to spot, here the force, is better recognised than the most latent one, here the level (Heseltine & Clemm von Hohenberg, 2024). The most frequent error is identified: lines uttered in character are coded at the fiction level, so that the share of in-character lines is underestimated. The codebook, the example and the model's outputs are kept, as the use of a proprietary model, liable to change without notice, requires (Ollion, Shen, Macanovic, & Chatelain, 2024).

4.3. The errors of the chain

Each step so far produces errors: words mistranscribed or invented, segments attributed to the wrong voice, gender misestimated, acts miscoded. We assume that they are distributed at random in the corpus, that is, that they depend neither on the system, nor on the GM, nor on the number of players, nor on the share of women. Under this assumption, errors blur the measurement of each table without digging any difference between tables: they make a real effect harder to detect, they do not make one appear that does not exist. A systematic error that affects all tables in the same way, like the coding of in-character lines as fiction, changes the description of the corpus without changing the comparisons between tables. What would put the results at risk is an error linked to a factor, for instance an annotator less accurate on one system than on another; this is what the assumption rules out (Section 10.6).

Table 2. The twenty kinds of act, five forces crossed with four levels. In each cell, the share of that kind of act among the acts of the annotated corpus, then a real example from the GM and one from the players (P), all drawn from a single session (Mugen-entertainment, *Chariot of the Gods*), lightly rephrased and translated from the French. Totals by force are in the first column, by level in the header.

force ↓ · level →	fiction (65%)	character (9%)	game (15%)	out-of-game (11%)
attempts (6%) · play a move	5% P "I'm going to open the pod with the abomination in it."	< 1% P "I yell: hurry up!"	1% P "I rolled Observation, I failed."	< 1%
asserts (60%) · supply content	42% GM "The first door opens, you come into an airlock." P "She's alive, yes."	5% GM "Me, I've got a service pistol." P "Captain, I torched the creature's nest."	8% GM "I'll put the stats back up, you just have to copy them." P "I'm already overloaded."	6% GM "Well, you're muted." P "Your mic is on!"
validates (5%) · accept or refuse	3% GM "Fine, you can join them, no problem." P "Yes, all right, works for me."	1% GM "Impossible, Captain, the order is written into my circuits."	1% GM "Yes, you can roll Command."	< 1% GM "Oh, if you want, we can."
asks (21%) · request	12% GM "What do you do?" P "How many arms do they have?"	3% GM "Hi Cham, you all right?" P "Come back quickly, Wilson, you have no weapon."	4% GM "Roll me an Observation check." P "Can I run a diagnostic?"	2% GM "Are you managing or not?" P "Can you hear me now?"
expresses (8%) · react	3% GM "Ah, that's horrible!" P "Damn, we're going to have to carry him!"	1% P "What the hell is this?"	1% GM "Oh damn, well played, two successes!"	3% GM "Well, good evening everyone." P "I think the alien spat on your mic."

4.4. Result: the acts

Table 2 gives the share of each of the twenty kinds of act in the annotated corpus. Two thirds of the acts bear on the fiction, one in seven on the game, one in ten on the life of the table, one in eleven is uttered in character. Three acts in five are *asserts*, one in five *asks*; *attempts*, *validates* and *expresses* each make between 5 and 8%. The GM produces 38% of the acts for more than half of the speaking time (Section 3.4): the GM's interventions are longer than the players', as expected of the one who narrates. Bales's (1950) scheme already gave, for group discussion, a majority of acts that supply content and a minority of questions and reactions; role-playing adds the game level and the in-character lines. Table 3 shows three minutes of a session as the annotation renders them.

5. From acts to transitions

5.1. Sequences and states

Each session becomes a sequence of acts in the order in which they were produced, each act reduced to three pieces of information: who speaks (the GM or a player, P), the force and the level. "GM asserts-fiction" or "P asks-game" are **states**. The *none* and *unreadable* turns are removed, the *continuation* turns merged into the act they extend, and two consecutive acts by the same participant bearing the same code make a single one, a narration cut into several turns counting only once. Twenty kinds of act for two roles make 40 possible states; 39 appear at least 20 times in the corpus.

Thirty-nine states are too many to measure within each table. They were reduced to 18 by merging, step by step, the two states that most resemble each other in what precedes and what follows

them, up to the point where going further would lose too much information about the sequences (Appendix C); a GM act is never merged with a player act. The merges are regular (Table C1). Out of game and in character, all the forces of a given role merge together, with one exception (the players' *expresses* out of game); at the game level, they group in twos and threes; in the fiction, they remain distinct, with two rare exceptions (the players' *validates* joins their *asserts*, the GM's *attempts* joins the GM's *expresses*). Merged states bear a "+": "P asserts+fiction" denotes the players' *asserts* and the *validates* merged into it. Sixteen of the 18 states make at least 1% of the acts; they are the ones the figures draw.

5.2. Result: the transition graph

The first description of a sequence is its **transition matrix**: for each state, the share of each state among those that immediately follow it, which we call a **transition proportion**. Figure 2 draws it for the whole corpus, keeping only the transitions that make up at least 10% of the departures from a state; the whole matrix, row by row, is in Appendix C (Figure C1). The figure shows three regularities.

First, each level closes on itself through an alternation between the GM and the players. In the fiction, the GM's *asserts* is followed by a player's *asserts* in 41% of cases, the player's *asserts* by a GM *asserts* in 29%, and a player's question by a GM *asserts* in 44%. In character, the GM's line is followed by a player's line in 62% of cases; at the game level and out of game, GM and players alternate in 30 to 38% of cases. The tendency to return to the previous speaker, familiar from models of turn-taking in small groups (Parker, 1988), takes here the form of a back-and-forth between one role and the others.

Table 3. Three minutes of a session (Mugen-entertainment, *Chariot of the Gods*, turns 540 to 556) as the automatic annotation renders them; utterances translated from the French. A turn can contain several acts (554, 556). Turns 542 to 544 continue the description begun in 541 and do not count as acts; nor does the "Okay" of turn 549. The right-hand column draws the sequence as the rest of the analysis sees it: one bubble per act, bearing the speaker, the force and the level, blue for the GM and orange for a player as in the figures, linked to the next act by an arrow; turns that are not acts have no bubble, the sequence passes through them.

turn	time	who	act	what is said	sequence
540	1:20:31	P1	attempts-fiction	I'm going to the cryostasis area.	P1 attempts fiction
541	1:20:36	GM	asserts-fiction	So there's an airlock, there are three other areas, but it's glazed, so you'll be able to see fairly easily from one area to the next...	GM asserts fiction
542 to 544	1:20:57	GM	continuation	... some are filled; when you count them, five or six have corpses in them...	
545	1:21:44	GM	asserts-out-of-game	I can't hear you, Tommy.	GM asserts out-of-game
547	1:21:50	P1	asks-fiction	The whole tank, can we see the whole tank, or is part of it hidden?	P1 asks fiction
548	1:21:54	GM	asserts-fiction	You can see almost all of it.	GM asserts fiction
549	1:21:57	P1	none	Okay.	
550	1:21:58	P1	asserts-fiction	Well, then, it might be worth checking whether the other vital signs really are genuine readings...	P1 asserts fiction
551	1:22:08	GM	asserts-fiction	You can tell it's been tampered with: on the outside of the cryotube there's a panel that's open...	GM asserts fiction
552	1:22:36	P1	asks-fiction	Do we have androids on our own ship?	P1 asks fiction
553	1:22:50	GM	asserts-fiction	No, as far as you know you don't, and as far as you know you have no idea.	GM asserts fiction
554	1:22:56	P1	asks-character	Wilson, if you could ask Mother whether they had androids on their ship.	P1 asks character
554	1:22:56	P1	attempts-fiction	I'm going to go and check the cryotubes.	P1 attempts fiction
555	1:23:04	P5	attempts-fiction	I'll ask.	P5 attempts fiction
556	1:23:08	GM	validates-fiction	Yes, that's fine.	GM validates fiction
556	1:23:08	GM	asserts-fiction	Mother tells you that an android is indeed registered...	GM asserts fiction

Second, what leaves a level goes to the fiction, and first of all to its two most frequent acts, the GM's *asserts* (18% of all acts) and the players' *asserts* (25%): in Figure C1, theirs are the two dark columns. The fiction laid down by the GM and taken up by the players is the pivot around which everything turns, a measurable form of the narration described by Goodwin (1984), told by one and shaped by the participation of the others. The pair player's question, GM's answer, is the most regular adjacency pair (Schefflo, 2007) in the corpus.

Third, the acts group by level in the figure, although its positions were chosen solely to cross as few arrows as possible.

6. From transitions to loops

6.1. Finding the loops

In a real session, acts intermingle, and what repeats is not an exact sequence of acts but a mixture that is stable for a certain time. We therefore look for the loops as stretches of the sequence during which the acts are drawn from the same mixture. The tool is a hidden Markov model (Rabiner, 1989), already used to describe the structure of a dialogue from its acts (Stolcke et al., 2000). Its principle: the table is assumed to be at every moment in one of k **regimes**, each producing the acts according to its own mixture and lasting a certain time before giving way to another, and the model finds by itself, in the 202 sequences, the k mixtures and the transitions that best account for what is observed. The number of regimes is chosen by the Bayesian information criterion, which weighs fidelity to the data against the complexity of the model (Appendix C). Such criteria retain more regimes than can be interpreted when data are abundant (Pohle, Langrock, van Beest, & Schmidt, 2017); the regimes that share the same dominant level are therefore merged, and these merged regimes are what we call loops. Each act finally receives its most probable regime, and everything else is counted on this decoding.

6.2. Result: four loops

Over the 202 sequences, the criterion prefers the largest number of regimes tried, ten. Each has a dominant level that makes up 82

to 97% of its acts: six fiction regimes (differing in the GM's share and in the players' questions or actions), two game regimes (one carried by the GM, the other by the players), one character regime and one out-of-game regime. Merged by level, they give **four loops** (Figure 3): the **fiction loop** (66% of the acts, 13.7 acts in a row on average), the **game loop**, where rules and dice are handled (15%, 4.5 in a row), the **character loop**, the lines uttered in character (10%, 8.6), and the **out-of-game loop** (9%, 6.4). When the fiction loop ends, it goes to the game loop 61 times out of 100, out of game 20 times, to the character loop 19 times; and everything returns to the fiction, 88% from the game loop, 77% from the character loop, 74% from out of game. The game loop most often opens with the GM's call for a roll, the character loop with a GM character speaking, the out-of-game loop with a remark by the GM.

The sequences that practitioners describe are thus found as loops, with two differences: the progression of the fiction and the search for information form a single loop, the two intermingling too finely to be separated, and out-of-game talk, which descriptions of play ignore, is a loop on a par with the in-character lines. The fiction loop is a collaborative floor (Edelsky, 1981), built by several, whereas the game loop is a floor held in turn, opened by the GM. One thing the model cannot see is players' discussion of what they are going to do: nothing marks it in the annotation, the addressee of an act not being coded, whereas it is one of the dimensions that distinguish multiparty dialogue from two-party dialogue (Traum, 2004).

7. From loops to distances between tables

7.1. One matrix per table

For each table, all its sessions pooled, we count the transitions from one act to the next: a matrix of 18 rows and 18 columns, where each cell counts how many times the row state is followed by the column state. We count likewise the transitions from one decoded loop to the next, a 4×4 matrix, which also counts the

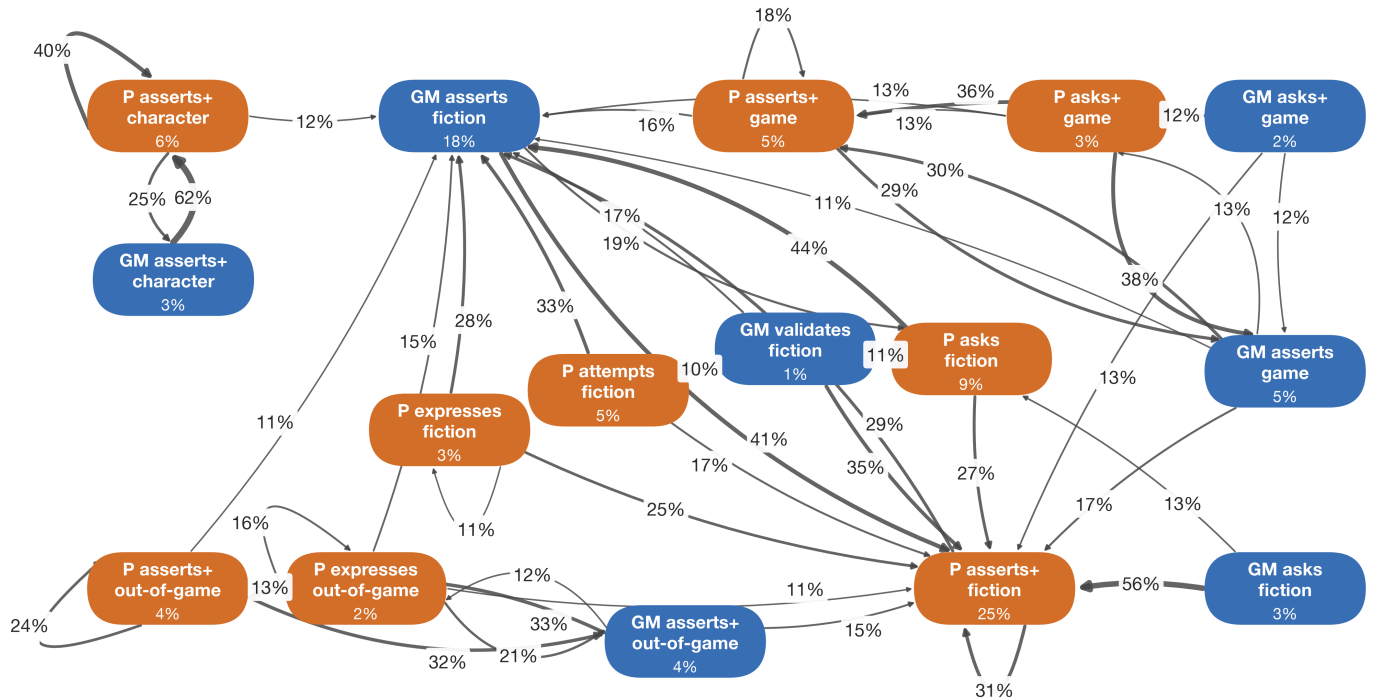


Figure 2. The graph of frequent transitions over the annotated corpus (18 merged states). A bubble is a state, written role, force and level: in blue a GM act, in orange a players' act; all bubbles have the same size, and the percentage in each is the state's share of the acts; only the 16 states that make up at least 1% of the acts are drawn. An arrow from a to b is the transition proportion from a to b, the share of the transitions that leave a and go to b; only those of at least 10% are drawn (46 arrows, which cover 63% of the transitions), the thickness being proportional. Positions are chosen so that the arrows cross as little as possible and carry no information in themselves. No model is involved: this is the transition matrix drawn.

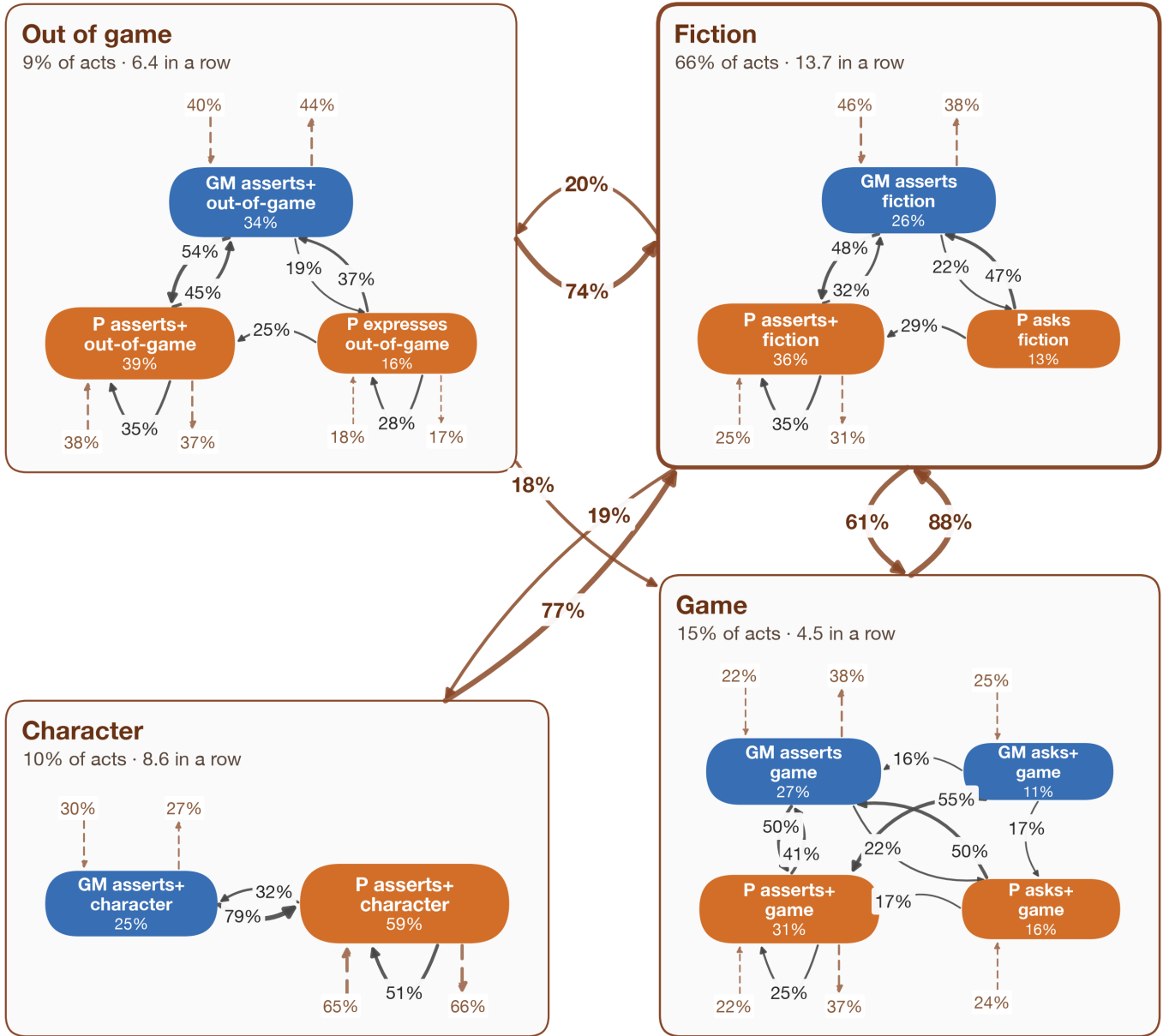


Figure 3. The four loops of the corpus, obtained by a hidden Markov model with ten regimes merged by dominant level. Each frame is a loop, named after its level; the thickness of its border is its share of the acts, and its subtitle gives that share and the mean duration of a passage (number of acts in a row). Within a frame, a bubble is an act (blue: the GM; orange: a player), its size and its percentage are its share of the loop's acts; only bubbles of at least 7% are drawn. A grey arrow between two bubbles is the share of the transitions from the first to the second among those that stay within the loop; a grey arrow returning to a bubble is the share of repetitions of that act by another participant. A brown arrow between two frames is, when the departure loop ends, the share of exits that go to the other loop; the brown arrows leaving a frame add up to 100%. A dotted brown arrow arriving from nowhere on a bubble is the share of entries into the loop that occur through that act; one leaving towards nowhere, the share of exits of which it is the last act. In all, 52 arrows are drawn.

times the loop continues. Only the transition proportions enter what follows: neither the frequency of the states nor the time spent enters, and the two **grains**, acts and loops, are taken together as a single problem. Transition proportions are measured in **points**: one point is one percentage point of a transition proportion, and the sizes of effects below are all in that unit.

7.2. The distance and its correction

Between two tables, at each grain, we measure how much their transition proportions differ, row by row, by a classical distance between probability distributions (the Jensen-Shannon divergence, Appendix D), each row weighted by the frequency of its state in the corpus.

This distance must be corrected. Two tables that played in exactly the same way would not be at distance zero, because each is observed only over a finite number of acts, and proportions estimated on few acts fluctuate: a short table looks far from all the others through the sole effect of chance. Before correction, the mean distance from a table to the others depended strongly on its

length (correlation -0.59 at the act grain, -0.30 at the loop grain), and chance made up 42% and 35% of the mean raw distance. The share of the distance expected from chance alone was therefore estimated for each pair, by cutting each table into two random halves and measuring the distance between them, then subtracted (Appendix D). After correction, the distance from a table to the others no longer depends on its length (correlation -0.04 and -0.02 ; Figure D1).

The two grains are finally combined into a single distance, the sum of the two corrected distances, each divided by its mean so that the two grains weigh the same, expressed in units of the mean pair ($= 1$). The 60 tables form 1,770 pairs, hence 1,770 distances, which say how differently any two tables play.

8. What distinguishes one table from another

8.1. A single model

To attribute the variation between tables to its sources, we explain the distance between two tables by what separates them: do they

have the same GM or not, the same system or not, by how much do their numbers of players and their shares of women differ? A regression on the 1,770 pairs gives the weight of each of these terms (Appendix E; Mantel, 1967; Legendre, Lapointe, & Casgrain, 1994; Lichstein, 2007).

These weights have a simple reading, that of a variance decomposition. Suppose that a table's way of playing is the sum of five parts: an effect of its GM, of its system, of its number of players, of its share of women, and a remainder specific to the table. Two tables then differ by the parts they do not share: by all five if nothing measured is common to them; by all the parts except the GM's if they are run by the same GM on different systems; only by their numbers of players, their shares of women and their remainders if they are run by the same GM on the same system. For the two numerical factors, the further apart two tables are, the larger that part of their difference.

The regression compares all these situations at once. The weight of the "different GMs" term is the part of the difference that disappears when the GM is the same, and likewise for the system. The weights of the two numerical terms are what one more player, or 25 more points of women, add to the difference between two tables. The constant term is what remains between two tables that share everything measured, the part specific to each table. This reading is in the line of dyadic variance decompositions and of the partition of variation on distances (Kenny & La Voie, 1984; McArdle & Anderson, 2001). The sum of these contributions is the between-table variance, and the share of each factor is its contribution divided by the sum.

One consequence of the design must be kept in mind: the 18 pairs of tables with the same GM (Table A1) are the only ones that inform the GM's contribution, all the other pairs being "different GMs" pairs, which is why its uncertainty will be the largest.

8.2. Inference and power

Inference then answers three questions (Appendices E and F). Does a factor weigh more than chance would? This is tested by shuffling its labels at random between tables, for instance reassigning the GMs to the tables, and recomputing its weight thousands of times: the p-value is the share of shuffles that give a weight at least as large as the observed one. For the GM, a second version shuffles the GMs only among the tables of a given system, which asks whether the GM matters once the system is held fixed. With what precision is each share estimated? The model is recomputed on subsets of tables drawn at random; the spread of the results, rescaled as befits subsamples, gives the 95% interval. Finally, what effect is the corpus able to detect? Into the real data, after erasing the real effect of the factor, an artificial effect of known size is injected, 8 points, either spread over all transitions or concentrated on the passage from the fiction to the game loop. How many times out of a hundred the test detects it is the power, which says what a non-significant result allows one to rule out. The analysis plan was written before the computations were run, with eight declared variants (Appendix G): the same model without the noise correction, on the longest tables only, on the classic games only, at each grain alone, and so on.

8.3. Result: the variance shares

Figure 4a gives the apportioning of the between-table variance of the transition proportions, acts and loops combined. The **GM** accounts for 34% of the between-table variance (95% interval from 11 to 58%), the **system** for 23% (11 to 34), the **number of players** for 6% (2 to 10), the **share of women** for 2% (0 to 8). The remaining 34% are **specific to the table** (10 to 58%): they distinguish two tables with the same GM on the same system, at equal number of players and share of women. The permutation tests give $p = 0.001$ for the GM (0.004 with the system held fixed), $p = 0.0001$ for the system, $p = 0.009$ for the number of players and $p = 0.15$ for the share of women (Table E1).

Figure 4b situates these results against what the corpus can detect (Appendix F; the unit of 8 differs from one panel to the next). For the system, the number of players and the GM, the observed effect is at or above the size that the corpus detects four times out of five, the GM's least clearly. For the share of women, the corpus would have detected, four times out of five, an effect about twice the size of the one observed: if the gender composition of the table has an effect on the form of the conversation, it is at most very small.

The difference between the GM's share and the system's, 12 points, has a standard error of 12 points: the order between the two is not settled. What is settled is that together they make up more than half of the between-table variance ($57\% \pm 12$).

8.4. The shares hold without the out-of-game talk

Published sessions are often edited, and what editing cuts for certain is the out-of-game talk; each channel cuts in its own way, and from 3% to 31% of a table's acts are out of game. If editing made the shares of Figure 4, this is where it would act. We therefore reran the whole analysis without any out-of-game act, as if every such passage had been cut (Appendix H). The shares hold: the GM goes from 34% to 42% of the between-table variance, the system from 23% to 30%, the table-specific share from 34% to 22%, the share of women stays at 2%, and only the number of players loses the part of its effect that passed through the out-of-game loop. The out-of-game talk is thus mostly variance specific to each table, and removing it sharpens the distances rather than creating them.

9. Where each factor acts

9.1. The same model, arrow by arrow

This section is exploratory: it takes the model of Section 8 apart, arrow by arrow, to see which factor moves which transition, and follows one case down to the passages themselves. We fitted the same model to each of the 52 arrows of Figure 3 taken on its own: for each arrow, its proportion is computed table by table, and the differences between tables in that one proportion are apportioned among the four factors exactly as in Section 8, with the same permutation tests (Appendix I). We say that a factor is **established** on an arrow when its permutation test gives $p \leq 0.05$. Figure 5 is the result: each arrow of Figure 3 is coloured by the factor established on it, with the size of the effect in points and, for the number of players, a sign that says whether the proportion rises or falls as players are added. Two cautions apply (Appendix I). A colour says that a factor is established on an arrow, not that it is the largest there, and the GM's test, which rests on 18 pairs of tables, has less power arrow by arrow than the others. And 52 arrows tested at the 5% level yield a few coloured arrows by chance alone: the tendencies, not any single arrow, are what counts.

9.2. Result: where the colours fall

Of the 52 arrows, 24 have at least one factor established and 28 have none. Three tendencies stand out.

The **number of players** acts where players hand over to one another. All its arrows leave a player's act or join the character loop, and its sign is always the one that turn-taking predicts: positive when the transition goes to another player (+4 to +7 points), negative when it goes to the GM (-3 to -6 points). Over the whole corpus, the share of player acts followed by a GM act falls by 6.9 points per additional player: the effect of the number of players is first of all this arithmetic.

The **system** acts around the game loop: on the passages that lead to it from the fiction (8 points) and from the out-of-game loop (6), on its opening by the GM's call for a roll (9) and on one exchange within it (7). On the entry into the game loop, the systems line up from *The Haunting*, which enters it least often, to *Lost Mine of Phandelver*, which enters it most, the "apocalypse"

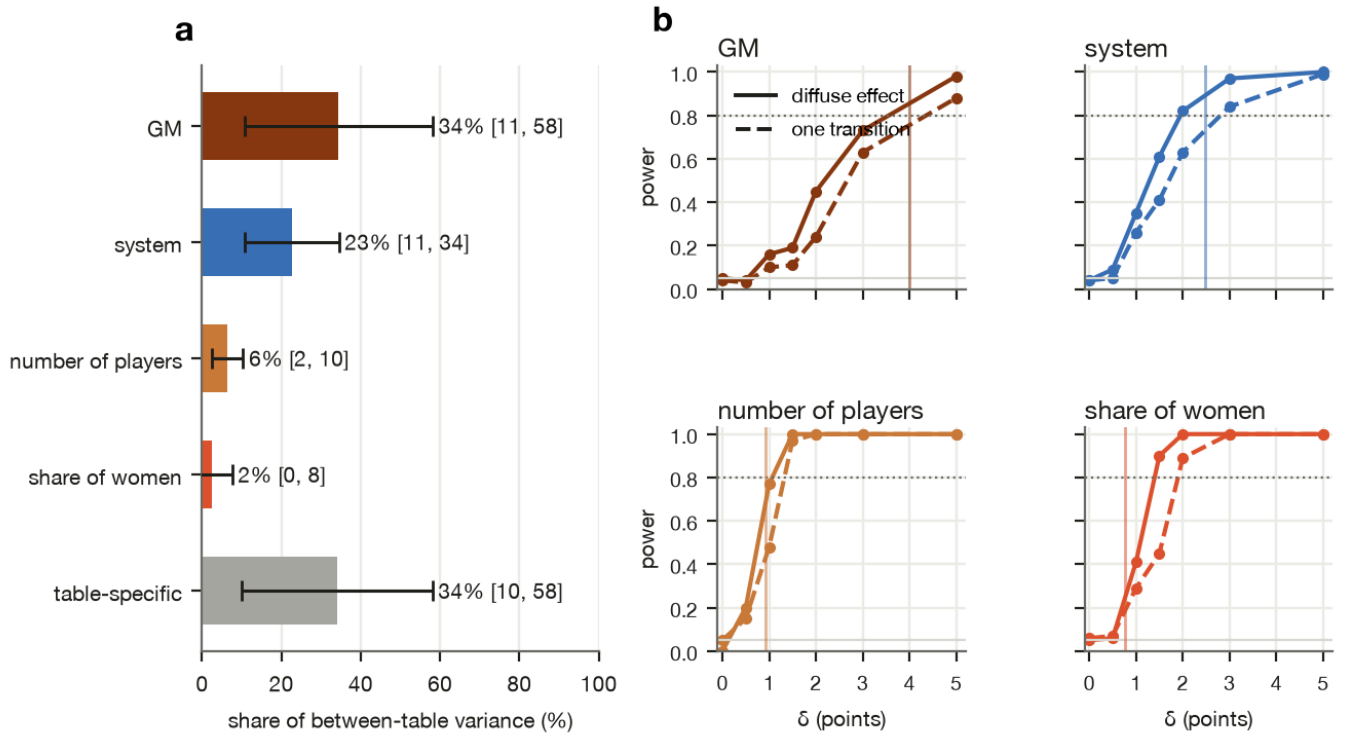


Figure 4. (a) The apportioning of the between-table variance of the transition proportions (acts and loops combined) among the four factors and what remains specific to the table, in a single model on the 1,770 pairs of 60 tables; error bars are the 95% intervals by rescaled subsampling of tables (Appendix E). (b) The power of the test of each term as a function of the size δ of the effect injected into the real matrices, in points of transition proportion (solid line: diffuse effect on all rows; dashes: a single transition shifted, fiction $\rightarrow$ game loop); for the system and the GM, δ is the standard deviation of the effect across categories, for the number of players the effect per player, for the share of women the effect per 25 points; the dotted line marks 80% power, the grey line at the bottom the 5% type I error rate, and the coloured vertical line the size of the observed effect translated into δ (diffuse form). One hundred replicates per size, 199 permutations per test.

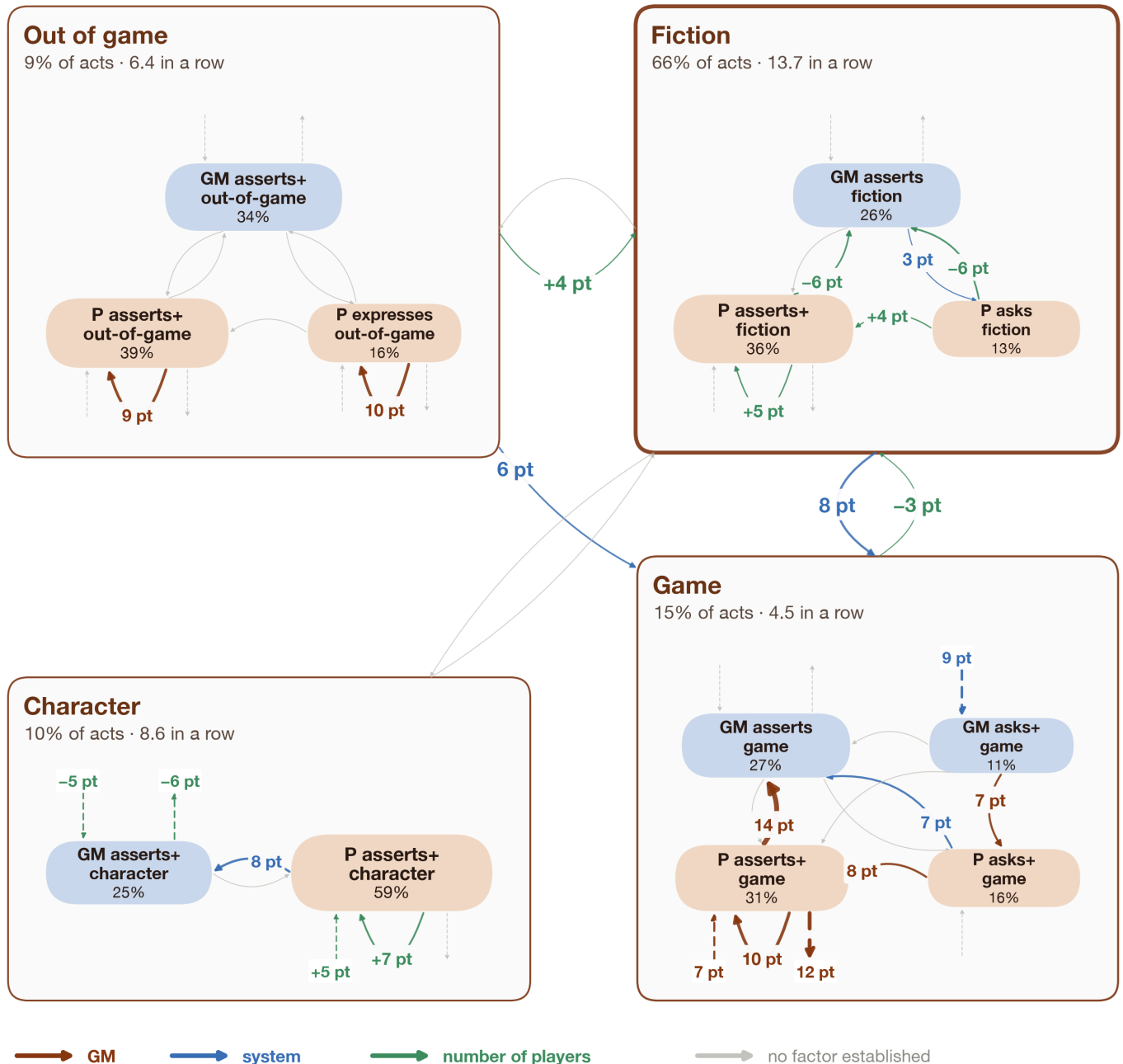


Figure 5. Figure 3 with each arrow coloured by the factor that moves its transition proportion most between the 60 tables, among the factors established on that arrow ($p \leq 0.05$ by permutation; brown: the GM; blue: the system; green: the number of players); grey: no factor established. The number on a coloured arrow is the standard deviation between tables of that transition proportion due to the factor, in points, the unit of δ in Figure 4; for the number of players, a + sign means that the proportion rises with the number of players and a - sign that it falls. The width of a coloured arrow is proportional to the variance explained by its factor; grey arrows are drawn thin whatever their transition proportion, which Figure 3 gives. Frames, bubbles and the 52 arrows drawn are those of Figure 3.

games in between (Appendix I). Outside the game loop, the system is established on a single arrow (8).

The **GM** carries the largest effects, 7 to 14 points, and they concentrate inside the game loop, on what follows a player's statement at the game level, a dice result or a modifier: whether the GM answers it, whether another player does, or whether it closes the passage. The GM is also established on the repetitions out of game.

9.3. A case: what the GM does inside the game loop

The colours of Figure 5 say where each factor acts, not what happens there. We take one case and follow it down to the passages themselves: the GM's effect inside the game loop, where its largest arrows lie.

Three measures were computed on the passages through the game loop (Appendix J). The **take-over rate** is how often the GM

answers a player's statement at the game level, a dice result or a modifier, with a statement of their own. It ranges from 17% to 93% depending on the table, and is about the same at the several tables of one GM whatever the system (70 to 78% at the tables of Les JDR du Gardien, 30 and 33% at those of bakadevil). The **narration length** is the number of words the GM spends in the fiction once the dice have spoken: from 10 to 82 words at the median, stable across the tables of one GM, and unrelated to the take-over rate (Spearman's $\rho = 0.05$). The **player-authored return** is the share of passages after which a player, not the GM, states the first fact of the fiction: from 12% to 52%. Who handles the mechanics and what their outcome becomes are thus two traits of the GM.

Reading passages from tables at the two ends of these measures suggested that GMs differ in what they make of a result once it is known. We therefore distinguished four **resolution styles** by the first statement about the world that follows the last result of a passage (Table J1). In the **narrator** style, that statement is the

GM's and adds at least one element that the outcome does not imply; in the **laconic** style, it is the GM's and restates the outcome bare; in the **accountant** style, it does not come, the result staying a game state; in the **delegated** style, it is a player's, and the GM lets it stand. A codebook fixes the unit, the boundaries and an ordered decision procedure, and records beside the style the type of roll, whether the stakes were stated before the roll and whether the outcome adds a consequence beyond success or failure (Appendix J). As a first calibration, 199 passages were drawn from 25 tables, eight per table and balanced by type of roll, and labelled by a language model (Claude; Anthropic, 2026), a single coder. The style belongs first to the GM: one GM is narrator in every passage over four tables and two systems, another in 95% of them over three tables and three systems, and only one changes markedly from one of their systems to the other. With the type of roll held fixed, reassigning the GMs at random within each system almost never reproduces the differences observed between them ($p \leq 0.011$ for the shares of narrator and accountant passages and for a graded score of the fiction produced). Reassigning the systems within a GM, or the numbers of players within a system, reproduces them easily (Appendix J). What the system does move shows beside the style, in the direction its rules prescribe: in *Blades in the Dark* the stakes are stated before the roll far more often than anywhere else, and the "apocalypse" games add a consequence beyond success or failure in most of their passages; neither contrast, however, is established once the GM is held fixed. One coder, 199 passages and a fragile boundary between the laconic and the narrator styles make this a calibration rather than a measurement.

10. Discussion

10.1. What the corpus shows

What distinguishes one table's transition proportions from another's comes first from who runs the game and what is being played, together for more than half of the variance, without their split being settled. Then comes, far behind but clearly, the number of players, whose effect is largely the arithmetic of turn-taking. The share of women does not count, within the limits of what the corpus can see. A third of the variance remains specific to the table: two tables with the same GM on the same system still differ by about two fifths of what two unrelated tables differ by (Appendix G). Upstream of these differences, the structure of the conversation is the same everywhere: a GM who holds more than half of the talk, four loops organised by level, and a pivot made of the fiction laid down by the GM and taken up by the players, to which every detour returns.

10.2. Back to the loops

The hidden Markov model recovers regimes organised by level, although the grouping of states into levels is not given to it. This sharpens the initial hypothesis: what repeats in a session is not an exact subsequence but a stable mixture of acts at the same level, and the sequences internal to each level are a GM $\rightleftharpoons$ players alternation, just as the phases of a group discussion are mixtures of acts rather than scripts (Bales & Strodtbeck, 1951). The search for information survives as a short motif, the question-answer pair, inside the fiction loop. The sequence that Table 3 shows at turns 552 to 556, a player asks, the GM asserts, a player attempts, the GM validates then asserts, is a variant of the classroom initiation-response-evaluation sequence (Sinclair & Coulthard, 1975; Mehan, 1979), except that the player initiates and the GM responds and evaluates. It is also the exchange that practitioners describe as the elementary unit of role-playing, a proposal followed by its acceptance or refusal (Féasson, 2016a, 2016b). Players' discussion of what they are going to do is not observable with this annotation (Section 6.2); the effect of the number of players on player $\rightarrow$ player sequences is probably its indirect trace.

10.3. The GM and the table

Where the same person runs several tables, these tables resemble each other act by act, whatever the system. If the GM holds the frame (Fine, 1983; Cover, 2010) and if the GM's skills carry over from one game to the next, as practitioners' handbooks assume (David & Larré, 2016), this is what one expects; editing, at least, does not make this resemblance (Section 8.4). The table-specific share of the variance, a third, is compatible with the idea that a group converges on its own way of playing (Fusaroli & Tylén, 2016) and that the relations between participants predict who speaks after whom (Gibson, 2003); the corpus measures neither.

Where the GM's effect is established transition by transition, it is inside the game loop (Section 9), and it takes two forms that do not go together: who handles the mechanics, and what the GM makes of a result once it is known. The resolution styles of Section 9.3 give the second a content. Same rulebook, same dice, same number: for the narrator, the dice make the scene; for the accountant, they keep the scene's accounts; for the GM who delegates, they give the player the right to write it. We put forward as a hypothesis that the coupling between mechanics and fiction at a table depends on this manner of the GM at least as much as on the rulebook, and that it is part of what the "GM" term of the model captures. It is also where the GM's effect would matter for what is played rather than for how the talk is organised. The first coding attributes the style to the GM and not to the system; it remains to extend it to the 2,000 passages of the corpus.

10.4. The system

The system can be read mostly in the frequency and duration of the detours through the game loop; the *Dungeons & Dragons* scenarios, richer in rolls and combat, are at one pole, *The Haunting* at the other. This is compatible with the idea that the system organises what each participant may say (Edwards, 1999; Young, 2005; Sintès, 2013) and that its rules preallocate part of the turns like those of a board game (Hofstetter, 2021), but in a modest version: the system does not restructure the conversation, it governs the frequency of one of its detours. In particular, the "apocalypse" games, designed so that the rules arbitrate the conversation (Baker, 2010, 2019), do not differ from the classic games in the GM's share of speech, and sit, on the entry into the game loop, between *The Haunting* and *Death House*. A system is also, in this corpus, an introductory scenario, a community of channels that plays it, a period of publication; and a "system" effect on distances may reflect a difference in dispersion between systems as much as a difference in location (Anderson & Walsh, 2013).

10.5. The number of players and the share of women

The effect of the number of players is the one predicted by the mechanics of turn-taking (Sacks et al., 1974): one more player, and a player's act is followed by the GM nearly 7 points less often. It joins the observation that the form of a discussion changes with its size (Fay et al., 2000) and that the first speaker dominates all the more as the group is larger (Bales, Strodtbeck, Mills, & Roseborough, 1951): here, the GM's share of speech falls with the number of players, but each player's share falls more. The absence of an effect of the share of women bears on the form of the sequences, not on the amount of talk per person nor on the content. A hobby can be gendered in who plays it and in what is said there (Fine, 1983; Dashiell, 2020) without the composition of the table changing the order in which acts follow one another, which is consistent with gender differences in talk that are small and context-dependent (Leaper & Ayres, 2007).

10.6. Limitations

The first limitation is the corpus, which is not a sample of role-playing tables in general. Published sessions are not ordinary tables: the recording puts the awareness of a further audience into the minds of the participants and, at the high end, turns a hobby session into a professional performance (Jones, 2024). What edit-

ing removes for certain, the out-of-game talk, does not make the results (Section 8.4); what a heavier editing would do, nothing allows us to simulate. The classic systems are the introductory scenarios of their games and the “apocalypse” campaigns are taken in their opening sessions: the corpus describes the beginning of a campaign, and what it says of the system it says of a game and a scenario together. Above all, games without a GM are absent by construction: the collection went through published scenarios written for a GM, the chain identifies the GM as the longest voice that narrates (Section 3.2), and every measure that follows is defined relative to that role. The system factor therefore varies only within the family of games organised around a GM, one game system among others (Briand, 2014); sharing the GM’s authority over the narration, the design choice to which Briand attributes decisive consequences, is the one the chain cannot measure, and games such as *Fiasco* or *Microscope* would require another chain. The conclusion that the system does not restructure the conversation holds within that family and says nothing beyond it.

The second limitation is that system and GM are nested in this corpus: 41 GMs out of 48 have a single table. On the 19 tables of the 7 GMs with several tables, each of the two effects shows with the other held fixed, and they add up with no visible interaction (Appendix G). But their respective weight in the whole corpus rests on the assumption that these 7 GMs are representative of the others, which is not assured. A GM with several tables is almost always a single channel, often a single crew and a single set-up, so that the GM term of the model cannot be told apart from the channel that publishes the tables (Appendix G). To settle the matter, one would need a crossed corpus, in which GMs run several systems and each system is run by several GMs (Judd, Westfall, & Kenny, 2017). By simulation on the real matrices (Appendix F), a full crossing of about 7 GMs by 7 systems, some fifty tables, would detect an effect of 2 points on each side four times out of five, where the real design is at one time out of five; such a crossing does not yet exist anywhere, to our knowledge.

The other limitations are those of the method. The errors of the chain are assumed independent of the factors (Section 4.3); the GM, the number of players and gender being measured by voice, with known and uncorrected errors, the model is somewhat overconfident about these terms (Fong & Tyler, 2021). The system’s share of the variance depends on the convention that gives the same weight to the two grains: it is 10% at the act grain and 35% at the loop grain, whereas the GM’s is 34% at both (Appendix G). Finally, all of this describes the form of the conversation, not what is said in it.

10.7. Conclusion

Does system matter, in conversation? It does, for about a quarter of what distinguishes one table from another, and mostly by setting how often the talk enters the rules. The person who runs the game matters at least as much, in a way that the corpus cannot separate from the channel that publishes them, and a third of the differences belong to each table. The number of players moves the alternation of speakers; the share of women moves nothing measurable. Beneath these differences, the conversation has the same structure everywhere, organised by level around the fiction that the GM lays down and the players take up. The method, from the published video to the apportioning of variance, is documented end to end and rests on public recordings; every figure and table of this article can be regenerated from the kept codebook, annotations and tools, up to the annotator (Appendix K).

References

Anderson, M. J., & Walsh, D. C. I. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecological Monographs*, 83(4), 557–574.

Anderson, T. W., & Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics*, 28(1), 89–110.

Anthropic (2026). *Claude Opus 5* [language model]. <https://www.anthropic.com>

Aronow, P. M., Samii, C., & Assenova, V. A. (2015). Cluster-robust variance estimation for dyadic data. *Political Analysis*, 23(4), 564–577.

Asher, N., Hunter, J., Morey, M., Benamara, F., & Afantenos, S. (2016). Discourse structure and dialogue acts in multiparty dialogue: The STAC corpus. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, 2721–2727.

Austin, J. L. (1962). *How to do things with words*. Oxford: Clarendon Press.

Baker, D. V. (2019). Powered by the Apocalypse, part 1. *lumpy games* [blog post]. <https://lumpy.games/2019/12/30/powered-by-the-apocalypse-part-1/>

Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups*. Cambridge, MA: Addison-Wesley.

Bales, R. F., & Strodtbeck, F. L. (1951). Phases in group problem-solving. *Journal of Abnormal and Social Psychology*, 46(4), 485–495.

Bales, R. F., Strodtbeck, F. L., Mills, T. M., & Roseborough, M. E. (1951). Channels of communication in small groups. *American Sociological Review*, 16(4), 461–468.

Boritchev, M., & Amblard, M. (2022). A multi-party dialogue resource in French. *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022)*, 814–823.

Bredin, H. (2023). pyannote.audio 2.1 speaker diarization pipeline: Principle, benchmark, and recipe. *Proceedings of Interspeech 2023*, 1983–1987.

Briand, R. (2014). *Le Maelstrom*. Self-published.

Bunt, H., Alexandersson, J., Choe, J.-W., Fang, A. C., Hasida, K., Petukhova, V., Popescu-Belis, A., & Traum, D. (2012). ISO 24617-2: A semantically-based standard for dialogue annotation. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, 430–437.

Burkhardt, F., Wagner, J., Wierstorf, H., Eyben, F., & Schuller, B. (2023). Speech-based age and gender prediction with transformers. *arXiv*, 2306.16962.

Callison-Burch, C., Tomar, G. S., Martin, L. J., Ippolito, D., Bailis, S., & Reitter, D. (2022). Dungeons and Dragons as a dialog challenge for artificial intelligence. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 9379–9393.

Chalk, A. (2023). Mapping an online production network: The field of “actual play” media. *Convergence*, 29(2), 449–466.

Chen, Y., Zheng, S., Wang, H., Cheng, L., Zhu, T., Song, C., Huang, R., Ma, Z., Chen, Q., Zhang, S., & Chen, X. (2024). ERes2NetV2: Boosting short-duration speaker verification performance with computational efficiency. *arXiv*, 2406.02167.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Core, M. G., & Allen, J. F. (1997). Coding dialogs with the DAMSL annotation scheme. *Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*, 28–35.

Cover, J. G. (2010). *The creation of narrative in tabletop role-playing games*. Jefferson, NC: McFarland.

Dashiell, S. L. (2020). Hooligans at the table: The concept of male preserves in tabletop role-playing games. *International Journal of Role-Playing*, 10, 26–39.

David, C., & Larré, J. (2016). Introduction : décider de sa pratique. In C. David & J. Larré (Eds.), *Mener des parties de jeu de rôle* (pp. 7–15). Saint-Orens-de-Gameville: Lapin Marteau.

DeBruine, L. M., & Barr, D. J. (2021). Understanding mixed-effects models through data simulation. *Advances in Methods and Practices in Psychological Science*, 4(1), 1–15.

Deng, K., Mehta, P. G., & Meyn, S. P. (2011). Optimal Kullback-Leibler aggregation via spectral theory of Markov chains. *IEEE Transactions on Automatic Control*, 56(12), 2793–2808.

Edelsky, C. (1981). Who’s got the floor? *Language in Society*, 10(3), 383–421.

Edwards, R. (1999). System does matter. *The Forge* [online]. http://www.indie-rpgs.com/_articles/system_does_matter.html

Endres, D. M., & Schindelin, J. E. (2003). A new metric for probability distributions. *IEEE Transactions on Information Theory*, 49(7), 1858–1860.

Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological Science*, 11(6), 481–486.

Féasson, V. (2016a). Pour une grammaire du jeu de rôle 3 : propositions et validation. *L’Averse* [blog post].

- <https://www.vivienfeasson.com/2016/04/27/pour-une-grammaire-du-jeu-de-role-3-echanges-elementaires-et-validation/>
- Féasson, V. (2016b). Pour une grammaire du jeu de rôle 4 : les gestes à la loupe. *L'Averse* [blog post]. <http://www.vivienfeasson.com/2016/07/13/pour-une-grammaire-du-jeu-de-role-4-les-gestes-a-la-loupe/>
- Fine, G. A. (1983). *Shared fantasy: Role-playing games as social worlds*. Chicago: University of Chicago Press.
- Fong, C., & Tyler, M. (2021). Machine learning predictions as regression covariates. *Political Analysis*, 29(4), 467–484.
- Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive Science*, 40(1), 145–171.
- Gibson, D. R. (2003). Participation shifts: Order and differentiation in group conversation. *Social Forces*, 81(4), 1335–1380.
- Giffin, H. (1984). The coordination of meaning in the creation of a shared make-believe reality. In I. Bretherton (Ed.), *Symbolic play: The development of social understanding* (pp. 73–100). Orlando, FL: Academic Press.
- Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30), e2305016120.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Cambridge, MA: Harvard University Press.
- Goodwin, C. (1984). Notes on story structure and the organization of participation. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 225–246). Cambridge: Cambridge University Press.
- Grosche, I., Bernaola-Galván, P., Carpena, P., Román-Roldán, R., Oliver, J., & Stanley, H. E. (2002). Analysis of symbolic sequences using the Jensen-Shannon divergence. *Physical Review E*, 65(4), 041905.
- Guillot, G., & Rousset, F. (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution*, 4(4), 336–344.
- Hendricks, S. Q. (2006). Incorporative discourse strategies in tabletop fantasy role-playing gaming. In J. P. Williams, S. Q. Hendricks, & W. K. Winkler (Eds.), *Gaming as culture: Essays on reality, identity and experience in fantasy games* (pp. 39–56). Jefferson, NC: McFarland.
- Heseltine, M., & Clemm von Hohenberg, B. (2024). Large language models as a substitute for human experts in annotating political text. *Research & Politics*, 11(1), 1–9.
- Hofstetter, E. (2021). Achieving preallocation: Turn transition practices in board games. *Discourse Processes*, 58(2), 113–133.
- Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71, 1–15.
- James, D., & Drach, J. (1993). Understanding gender differences in amount of talk: A critical review of research. In D. Tannen (Ed.), *Gender and conversational interaction* (pp. 281–312). New York: Oxford University Press.
- Jones, S. (2024). Documented and actual play in role-playing games. In J. P. Zagal & S. Deterding (Eds.), *The Routledge handbook of role-playing game studies* (Chapter 18, pp. 352–364). Abingdon: Routledge.
- Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, 68, 601–625.
- Jurafsky, D., Shriberg, E., & Biasca, D. (1997). *Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual, draft 13*. Boulder, CO: University of Colorado, Institute of Cognitive Science.
- Karpowitz, C. F., & Mendelberg, T. (2014). *The silent sex: Gender, deliberation, and institutions*. Princeton, NJ: Princeton University Press.
- Kenny, D. A., & La Voie, L. (1984). The Social Relations Model. *Advances in Experimental Social Psychology*, 18, 141–182.
- Kim, J. H. (1998). *The Threefold Model FAQ* [online]. https://darkshire.net/~jhkim/rpg/theory/threefold/faq_v1.html
- Koenecke, A., Choi, A. S. G., Mei, K. X., Schellmann, H., & Sloane, M. (2024). Careless Whisper: Speech-to-text hallucination harms. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1672–1681.
- Koudenburg, N., Postmes, T., & Gordijn, E. H. (2017). Beyond content of conversation: The role of conversational form in the emergence and regulation of social structure. *Personality and Social Psychology Review*, 21(1), 50–71.
- Leaper, C., & Ayres, M. M. (2007). A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech. *Personality and Social Psychology Review*, 11(4), 328–363.
- Legendre, P., Lapointe, F.-J., & Casgrain, P. (1994). Modeling brain evolution from behavior: A permutational regression approach. *Evolution*, 48(5), 1487–1499.
- Lerner, G. H. (2003). Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society*, 32(2), 177–201.
- Lichstein, J. W. (2007). Multiple regression on distance matrices: A multivariate spatial analysis tool. *Plant Ecology*, 188(2), 117–131.
- Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1), 145–151.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2), 209–220.
- McArdle, B. H., & Anderson, M. J. (2001). Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology*, 82(1), 290–297.
- Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard University Press.
- Miller, G. A. (1955). Note on the bias of information estimates. In H. Quastler (Ed.), *Information theory in psychology: Problems and methods* (pp. 95–100). Glencoe, IL: Free Press.
- Ollion, É., Shen, R., Macanovic, A., & Chatelain, A. (2024). The dangers of using proprietary LLMs for research. *Nature Machine Intelligence*, 6(1), 4–5.
- OpenAI (2026). *GPT-5.4 mini* [language model]. <https://openai.com>
- Ostyakova, L., Smilga, V., Petukhova, K., Molchanova, M., & Kornev, D. (2023). ChatGPT vs. crowdsourcing vs. experts: Annotating open-domain conversations with speech functions. *Proceedings of the 24th Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2023)*, 242–254.
- Parker, K. C. H. (1988). Speaking turns in small group interaction: A context-sensitive event sequence model. *Journal of Personality and Social Psychology*, 54(6), 965–971.
- Plaquet, A., & Bredin, H. (2023). Powerset multi-class cross entropy loss for neural speaker diarization. *Proceedings of Interspeech 2023*, 3222–3226.
- Pohle, J., Langrock, R., van Beest, F. M., & Schmidt, N. M. (2017). Selecting the number of states in hidden Markov models: Pragmatic solutions illustrated using animal movement. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3), 270–293.
- Politis, D. N., Romano, J. P., & Wolf, M. (1999). *Subsampling*. New York: Springer.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286.
- Rached, Z., Alajaji, F., & Campbell, L. L. (2004). The Kullback-Leibler divergence rate between Markov sources. *IEEE Transactions on Information Theory*, 50(5), 917–921.
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. *Proceedings of the 40th International Conference on Machine Learning, PMLR 202*, 28492–28518.
- Rameshkumar, R., & Bailey, P. (2020). Storytelling with dialogue: A Critical Role Dungeons and Dragons dataset. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5121–5134.
- Remme, L., & Tang, K. (2025). Playing with voices: Tabletop role-playing game recordings as a diarization challenge. *Findings of the Association for Computational Linguistics: NAACL 2025*, 4984–4998.
- Ritter, A., Cherry, C., & Dolan, B. (2010). Unsupervised modeling of Twitter conversations. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, 172–180.
- Roux, U. (2020). Actual plays : quand le jeu de rôle sur table se médiatise. *Recherches en communication*, 49, 133–152.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735.
- Salamini, H., & Vinciarelli, A. (2012). Automatic role recognition in multi-party conversations: An approach based on turn organization, prosody, and conditional random fields. *IEEE Transactions on Multimedia*, 14(2), 338–345.
- Sawyer, R. K. (1997). *Pretend play as improvisation: Conversation in the preschool classroom*. Mahwah, NJ: Lawrence Erlbaum.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, 5(1), 1–23.

- Searle, J. R., & Vanderveken, D. (1985). *Foundations of illocutionary logic*. Cambridge: Cambridge University Press.
- Shao, J., & Wu, C. F. J. (1989). A general theory for jackknife variance estimation. *The Annals of Statistics*, 17(3), 1176–1197.
- Sharp, A. J. (2023). *Conversational analysis of turn-taking in a Dungeons & Dragons game session* (master's thesis). San Francisco State University.
- Sinclair, J., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. London: Oxford University Press.
- Sintes, F. (2013). Le système de jeu de rôle. *Limbic Systems* [blog post]. <https://www.limbicsystems.jdr.com/le-systeme-de-jeu-de-role/>
- Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, 35(4), 627–632.
- Stephan, F. E., & Mishler, E. G. (1952). The distribution of participation in small groups: An exponential approximation. *American Sociological Review*, 17(5), 598–608.
- Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C., & Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3), 339–373.
- Torner, E. (2018). RPG theorizing by designers and players. In J. P. Zagal & S. Deterding (Eds.), *Role-playing game studies: Transmedia foundations* (pp. 191–212). New York: Routledge.
- Traum, D. R. (2004). Issues in multiparty dialogues. In F. Dignum (Ed.), *Advances in agent communication* (Lecture Notes in Computer Science 2922, pp. 201–211). Berlin: Springer.
- Traum, D. R., & Hinkelman, E. A. (1992). Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3), 575–599.
- Tychsen, A., Hitchens, M., Brolund, T., & Kavakli, M. (2005). The game master. *Proceedings of the Second Australasian Conference on Interactive Entertainment*, 215–222.
- Tychsen, A., Smith, J. H., Hitchens, M., & Tosca, S. (2006). Communication in multi-player role playing games: The effect of medium. In S. Göbel, R. Malkewitz, & I. Iurgel (Eds.), *Technologies for interactive digital storytelling and entertainment (TIDSE 2006)* (Lecture Notes in Computer Science 4326, pp. 277–288). Berlin: Springer.
- Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2), 260–269.
- Young, M. J. (2005). Theory 101: System and the shared imagined space. *Places to Go, People to Be*, 26 [online]. <http://www.ptgptb.org/0026/theory101-01.html>
- Zhu, A., Aggarwal, K., Feng, A., Martin, L. J., & Callison-Burch, C. (2023). FIREBALL: A dataset of Dungeons and Dragons actual-play with structured game state information. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 4171–4193.
- Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2024). Can large language models transform computational social science? *Computational Linguistics*, 50(1), 237–291.

Ludography

- Baker, D. V. (2010). *Apocalypse World*. Lumpley Games.
- Cameron, H. (2016). *The Sprawl*. Ardens Ludere.
- Chaosium (1981). The haunting. In *Call of Cthulhu*. French translation “La Maison Corbitt” in *L'Appel de Cthulhu*, 7th edition (Sans-Détour, 2015).
- Conway, B. (2016). *Masks: A new generation*. Magpie Games.
- Darrington Press (2023). *Candela Obscura*. Darrington Press.
- Darrington Press (2025). *Daggerheart*. Darrington Press.
- Free League Publishing (2019). *Alien: The roleplaying game and the scenario Chariot of the Gods*. French translation *Le Chariot des dieux* (Arkhané Asylum Publishing, 2020).
- Harper, J. (2017). *Blades in the Dark*. Evil Hat Productions.
- Magpie Games (2021). *Root: The tabletop roleplaying game*. Magpie Games.
- Magpie Games (2022). *Avatar Legends: The roleplaying game*. Magpie Games.
- Morningstar, J. (2009). *Fiasco*. Bully Pulpit Games.
- Robbins, B. (2011). *Microscope*. Lame Mage Productions.
- Sands, M. (2015). *Monster of the Week* (revised edition). Evil Hat Productions.
- Wizards of the Coast (2014). *Dungeons & Dragons starter set: Lost Mine of Phandelver*. French translation *La Mine perdue de Phandelver*.
- Wizards of the Coast (2016). *Curse of Strahd*, chapter “Death House”. French translation *La Malédiction de Strahd*.

Appendix A. From the recording to the speaking turns: voices, roles, gender, the GM’s share of speech and the list of GMs

A.1. Voices and roles. Voices were segmented by pyannote 3 (Bredin, 2023; Plaquet & Bredin, 2023), then represented by ERes2NetV2 speaker embeddings (Chen et al., 2024). That a participant’s role can be read in the very organisation of their turns is established for other conversations with instituted roles (Salamin & Vinciarelli, 2012); the GM is here the longest voice that narrates. On the tables added last, the player voices were taken as the clustering proposed them, then checked by the similarity between profiles and by a simple rule: the same person never overlaps with themselves, two groups that speak at the same time are two people. In all, 6 groups attributed to players were returned to the GM and 2 pairs of groups merged. Shared identities across tables were established by the similarity of the mean embeddings.

A.2. Voice gender. Three methods: the median fundamental frequency measured by Praat through Parselmouth (Jadoul, Thompson, & de Boer, 2018; threshold at 165 Hz), an age-gender classifier based on wav2vec 2.0 (Burkhardt, Wagner, Wierstorf, Eyben, & Schuller, 2023), and a two-component Gaussian mixture on the speaker embeddings. The gender retained is the mean of the three probabilities; the three methods agree for 92% of the voices, and 26 voices out of 325 remain flagged for checking.

A.3. The GM’s share of speech. Share of the time of the turns attributed to the GM in the time of all the turns of the table. Medians by system range from 49 to 65%. Effect of the system: Kruskal-Wallis test on the ten systems with at least two tables, $p = 0.70$; permutation test of F on the twelve systems, $p = 0.78$. Link with the number of players: Spearman’s $\rho = -0.50$, $p < 0.001$.

A.4. The GMs and their tables. Table A1 lists the 48 GMs identified by voice, with the tables each of them runs. Seven GMs run several tables; the other 41 run one. Eight channels have several GMs, one per campaign; conversely, each of the seven GMs with several tables runs a single channel.

Appendix B. The codebook

Table B1 gives the codes of the two dimensions and of the three non-acts, with the decision criterion of each, the order of the procedure and the main conventions.

Appendix C. From acts to states and loops

C.1. Reduction of the states. The 39 states (role $\times$ force $\times$ level appearing at least 20 times) were reduced to 18 by hierarchical aggregation with minimal loss of mutual information between an act and the next, a problem neighbouring the aggregation of Markov chains studied by Deng, Mehta and Meyn (2011), while forbidding the merging of a GM act with a player act; 95% of the original mutual information (0.655 bit) is retained. The move is the same as the one that brings the 220 label combinations of the Switchboard corpus down to 42 classes (Jurafsky, Shriberg, & Biasca, 1997), here with an explicit criterion. Table C1 lists the 18 merged states with the forces each of them absorbs and its share of the acts.

C.2. The transition matrix. Figure C1 gives the whole transition matrix of the corpus on the 18 merged states, of which Figure 2 draws the transitions of 10% and more between the 16 states of 1% and more.

C.3. The hidden Markov model. Model with categorical observations, fitted to the 202 sequences placed end to end, for k from 2 to 10 regimes with 6 initialisations each; the k retained is the one preferred by the Bayesian information criterion (Schwarz, 1978), which is here the largest tried, 10. Information criteria are known to retain more regimes than can be interpreted when data are

Table A1. The 48 GMs and their 60 tables, sorted by number of tables and read down the left half, then the right half. A GM is named by the channel that publishes their tables, with a letter when the channel has several GMs. Systems are those of Section 2.2, abbreviated: Haunting = *The Haunting*, Chariot = *Chariot of the Gods*, Phandelver = *Lost Mine of Phandelver*, Death House in *Dungeons & Dragons* unless marked Daggerheart, Blades = *Blades in the Dark*, MotW = *Monster of the Week*, Candela = *Candela Obscura*, Sprawl = *The Sprawl*, Avatar = *Avatar Legends*; (×2) marks a GM who runs the same system twice. Players are the player voices present in at least 2% of the turns, one figure per table. Pairs is the number of pairs of tables with the same GM, $n(n-1)/2$ for a GM with n tables: 18 in all, out of the 1,770 pairs of the corpus, and the only pairs that inform the GM's term of the model (Section 8.1).

GM	tables	systems	players	pairs	GM	tables	systems	players	pairs
Les JDR du Gardien	4	Chariot (×2), Haunting (×2)	4, 5, 4, 4	6	La Table Virtuelle (c)	1	Masks	4	
Jeux de Trolls	3	Chariot (×2), Haunting	5, 5, 3	3	Le Donjon d'Antares	1	Phandelver	4	
La Table Virtuelle (b)	3	Blades, MotW, Sprawl	3, 3, 3	3	Le Nain Gris JDR	1	Haunting	4	
Wasitum	3	Blades, Candela, Chariot	5, 4, 5	3	Le repaire du Rôliste	1	Chariot	6	
Abolas	2	Chariot, Death House	5, 4	1	Les Chroniques de R'lyeh	1	Haunting	2	
bakadevil	2	Chariot, MotW	6, 5	1	Malugatv	1	Haunting	4	
La Réponse Dé	2	Death House, Phandelver	6, 3	1	Mugen-entertainment (a)	1	Death House	4	
On3Up	1	Haunting	6		Mugen-entertainment (b)	1	Chariot	4	
Alea_jacta_sit	1	Death House	5		Naetherion	1	Death House	5	
Alex Thaumiel	1	Chariot	5		PolareBears	1	Chariot	4	
Antre de Guerre	1	Chariot	6		RPG Sul'Side (a)	1	Death House	5	
CaosFR	1	Phandelver	6		RPG Sul'Side (b)	1	Phandelver	4	
Donjons & Souillons	1	Phandelver	5		RPG Sul'Side (c)	1	Root	3	
Es-Tu Game? (a)	1	Death House	5		Rôliste Life	1	Haunting	3	
Es-Tu Game? (b)	1	Root	4		Sasmira	1	Phandelver	4	
Govin Jdr	1	Chariot	4		Studio 4D2 (a)	1	Death House (Daggerheart)	5	
Jeudis d'Outre-Manche	1	Phandelver	2		Studio 4D2 (b)	1	Chariot	5	
L'Abominable Gardien	1	Haunting	3		Studio 4D2 (c)	1	Candela	3	
L'armée des 12 rôlistes	1	Chariot	5		Studio 4D2 (d)	1	Avatar	5	
L_R0jo	1	Phandelver	5		The Midnight Lodge (a)	1	Chariot	5	
La Table de Rencontres (a)	1	Death House	5		The Midnight Lodge (b)	1	Chariot	4	
La Table de Rencontres (b)	1	MotW	4		Time To Roll (a)	1	Phandelver	4	
La Table de Rencontres (c)	1	Avatar	4		Time To Roll (b)	1	Chariot	4	
La Table Virtuelle (a)	1	Chariot	5		Time To Roll (c)	1	Sprawl	5	
				48 GMs					18
					60	12 systems			

Table B1. The codes. The order of the procedure is: for the force, asks, then validates, then attempts, then expresses, asserts by default; for the level, character, then out-of-game, then game, fiction by default. Authority: the GM has authority over the fiction and the rules, the player over their character and their sheet, which distinguishes “attempts” (the outcome does not depend on the speaker) from “asserts” (the speaker lays down what they have authority to lay down). Main conventions: an action named by an object of the rules is attempts-game; a dice result announced on its own is attempts-game; the result translated into the world is an asserts-fiction; validating while giving one's reason is a single act; a joke is coded at the level of its object; discourse markers are never annotated separately; commitments (“I'll do it”) are asserts; two consecutive turns by the same speaker bearing the same code make a single act.

dimension	value	definition	decision criterion
force	attempts	the speaker plays a move in the first person (the character acts, or the speaker performs an operation of the game) whose effect does not depend on the speaker	the utterance commits to an action whose outcome is to be decided by another or by the game
force	asserts	the speaker supplies content: description, information, opinion, narration, commitment	by default, when no other force applies
force	validates	the speaker accepts or refuses what another has proposed	there is a prior proposal by another participant
force	asks	the speaker requests information or an action	interrogative or imperative form, or recognisable indirect request
force	expresses	the speaker reacts, without laying down or asking anything	interjection, affective reaction, evaluation without new content
level	fiction	the world being played, spoken about as a participant	by default
level	character	words uttered live by a character	a character in the fiction could hear these words
level	game	the rules, the dice, the sheets, the virtual tabletop	the object of the act is an element of the game apparatus
level	out-of-game	the participants' life outside the game, technical matters, schedules	the object of the act is a real person or their environment
non-act	continuation	the turn continues the same speaker's previous act	same speaker, same code, content that extends
non-act	none	acknowledgement, echo	“okay”, “hmm”, repetition without addition
non-act	unreadable	unusable transcription	

Table C1. The 18 merged states. A merged state bears a “+” and is named after its dominant force; the third column lists all the forces of the given role and level that it gathers (“none” when the state is a single original state). Shares of acts over the annotated corpus (202 sessions).

level	merged state	forces merged into it	share of acts
fiction	GM asserts-fiction	none (asserts alone)	18.1%
fiction	GM validates-fiction	none (validates alone)	1.2%
fiction	GM asks-fiction	none (asks alone)	3.3%
fiction	GM expresses + ·fiction	expresses, attempts	< 1%
fiction	P attempts-fiction	none (attempts alone)	4.6%
fiction	P asserts + ·fiction	asserts, validates	25.4%
fiction	P asks-fiction	none (asks alone)	8.8%
fiction	P expresses-fiction	none (expresses alone)	2.9%
character	GM asserts + ·character	asserts, validates, asks, expresses	2.6%
character	P asserts + ·character	asserts, attempts, validates, asks, expresses	6.3%
game	GM asserts-game	none (asserts alone)	4.6%
game	GM validates + ·game	validates, expresses	< 1%
game	GM asks + ·game	asks, attempts	2.3%
game	P asserts + ·game	asserts, validates, expresses	4.8%
game	P asks + ·game	asks, attempts	2.8%
out-of-game	GM asserts + ·out-of-game	asserts, attempts, validates, asks, expresses	4.2%
out-of-game	P asserts + ·out-of-game	asserts, attempts, validates, asks	4.4%
out-of-game	P expresses-out-of-game	none (expresses alone)	2.4%

abundant, and Pohle et al. (2017) recommend pragmatic choices in that case; ours is to merge the regimes that share a dominant level. Each act receives the most probable regime by the Viterbi (1967) algorithm. A regime found without supervision mixes the structure and the content of what it groups (Ritter, Cherry, & Dolan, 2010); this is what is expected here, where the level is a property of the content.

Appendix D. The distance between tables and the correction of sampling noise

The distance. Between two tables, at each grain, the distance is the square root of the Jensen-Shannon divergence (Lin, 1991; Endres & Schindelin, 2003) between their transition rows, each row weighted by the frequency of its state in the whole corpus, the same weighting for all pairs, the rows being smoothed towards the common distribution of the corpus. It is the symmetric version of the divergence rate between two Markov sources, the sum of the divergences between rows weighted by the distribution of the states (Rached, Alajaji, & Campbell, 2004). Comparing Markov chains by their transition matrices has a long history (Anderson & Goodman, 1957); we take it through distances rather than through homogeneity tests, because the factors to be compared are mixed, categorical and continuous, and because the unit of inference is the table and not the transition. The two grains are combined into a multi-scale distance, $d^2 = d_{\text{acts}}^2 / \text{mean} + d_{\text{loops}}^2 / \text{mean}$.

The correction. Let P_{ik} be row k (the transition probabilities from state k) of the matrix of table i , estimated from n_{ik} transitions and smoothed towards the common distribution of the corpus with a weight $\lambda = 2$. The Jensen-Shannon divergence between the estimated rows of two tables is not an unbiased estimator of the divergence between their true rows: to second order, for two tables drawn from the same distribution, its expectation is about $(K - 1) / (8 \ln 2) \cdot (1/n_{ik} + 1/n_{jk})$ bits, K being the number of states, which is the two-sample special case of the approximate distribution of the estimator given by Grosse et al. (2002); this bias of information-based estimators has been known since Miller (1955). The squared distance between two tables therefore contains a noise term that decreases as the inverse of their sample sizes, and a short table is far from all the others through this effect alone.

We modelled this term as $\sum_k w_k c_k (x_{ik} + x_{jk})$, with $x = n / (n + \lambda)^2$ (the variance of a smoothed proportion) and w_k the frequency of state k in the corpus. The constants c_k were estimated empiri-

cally by cutting each table into two random halves (each transition drawn by coin toss, 20 times per table) and regressing the divergence between the two halves on $x_A + x_B$, row by row, by least squares through the origin; they match the theoretical value. The term is subtracted from each squared distance. The corrected distance can be slightly negative (3% of the pairs at the loop grain) since the noise is an expectation; this has no consequence for the model, which is linear in d^2 .

Without this correction, the variance components would be computed on distances of which nearly half is noise unrelated to the factors: the effects are halved and the GM, whose strongest evidence is the resemblance of four tables of which two are among the shortest in the corpus, ceases to be visible.

Appendix E. The single model: variance components and inference

The regression. For each of the 1,770 pairs of tables (i, j), the multiple regression on distance matrices (Mantel, 1967; Smouse, Long, & Sokal, 1986; Legendre et al., 1994; Lichstein, 2007) reads $d_{ij}^2 = c_0 + c_m \cdot [\text{different GMs}] + c_s \cdot [\text{different systems}] + c_j \cdot (\Delta \text{ players})^2 + c_f \cdot (\Delta \text{ share of women} / 0.25)^2 + \epsilon_{ij}$, the coefficients being estimated by least squares on all pairs at once, with predictors fixed before the analysis.

The model. Suppose that each table i has a position y_i in a Euclidean space, with $y_i = \mu + a_{m(i)} + b_{s(i)} + \beta_j u + \beta' f_i v + e_i$, where a is a random effect of the GM with total variance σ_m^2 , b an effect of the system with variance σ_s^2 , j and f the number of players and the share of women, u and v unit directions, and e a table-specific residual with variance σ_e^2 , all independent. Then the expectation of the squared distance between two tables is $E\|y_i - y_j\|^2 = 2\sigma_e^2 + 2\sigma_m^2 \cdot [m(i) \neq m(j)] + 2\sigma_s^2 \cdot [s(i) \neq s(j)] + \beta^2 (j_i - j_j)^2 + \beta'^2 (f_i - f_j)^2$. The regression of d_{ij}^2 on these five regressors therefore estimates $2\sigma_e^2$, $2\sigma_m^2$, $2\sigma_s^2$, β^2 and β'^2 ; the between-table variance attributable to a covariate is $\beta^2 \text{Var}(j)$. The Jensen-Shannon divergence is not exactly a squared Euclidean distance, but it is one to second order for close distributions, which is the case here; the Gower matrix of the corrected distances has 26 negative eigenvalues out of 60, for 9% of the positive trace, and an effective dimension of about eight.

The tests. Pairs are not independent, each table entering 59 of them (Aronow, Samii, & Assenova, 2015), and the Mantel test has been criticised for its type I error rate when units are not exchangeable (Guillot & Rousset, 2013); inference is therefore carried out at the level of tables. The p -value of a term is obtained

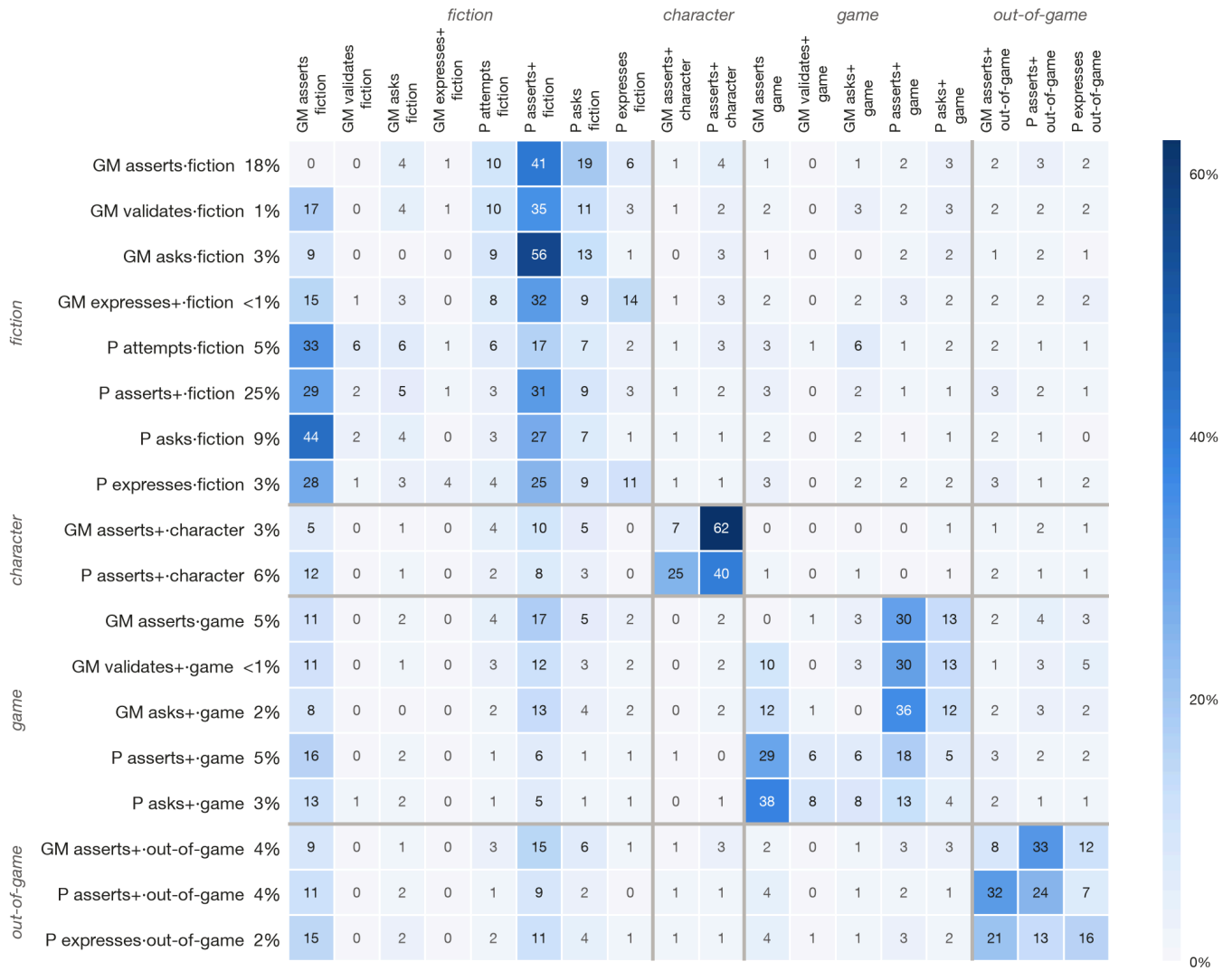


Figure C1. The transition matrix between the 18 merged states over the annotated corpus (125,058 transitions). A row is the state of an act, a column the state of the act that follows it, and a cell the transition proportion from the row state to the column state; each row adds up to 100%, and the diagonal is the repetition of the same act by another participant (a participant never repeats themselves, consecutive identical acts being merged). Rows and columns are ordered by level, then role (GM, then players), then force; the rules separate the levels; the share of each state among all acts follows its label; each cell carries its proportion rounded to the nearest per cent (0 standing for less than 0.5%), in grey below 5%. The blocks along the diagonal are the four levels closing on themselves; the two dark columns are the GM's and the players' assertions in the fiction, to which every level returns.

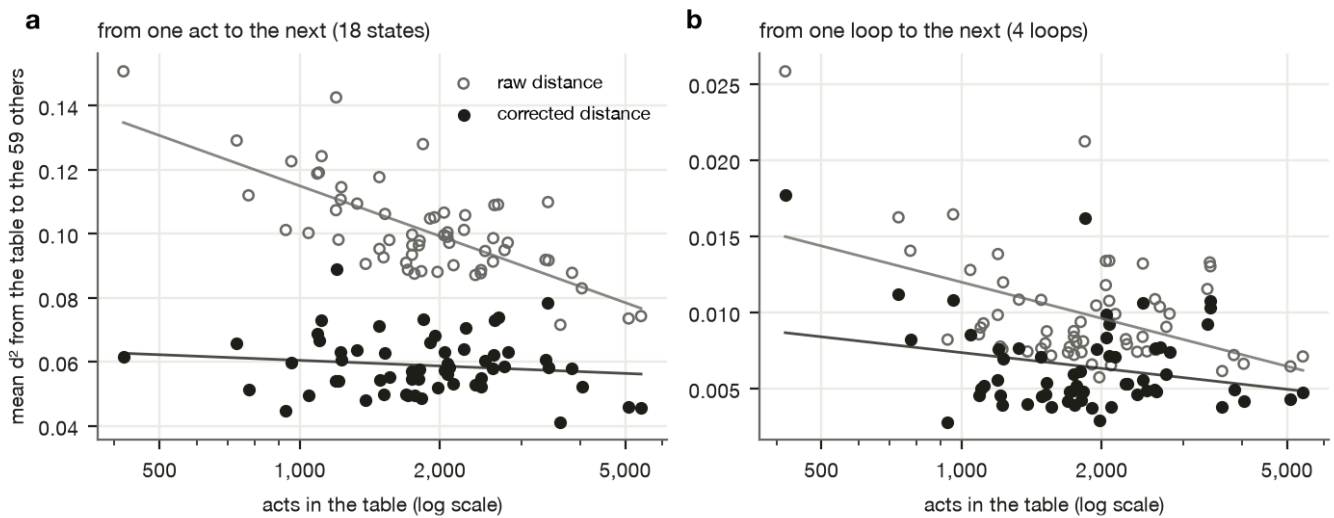


Figure D1. The correction of sampling noise. For each table, its mean squared distance to the 59 others as a function of the number of annotated acts it contains, before (open circles) and after (filled dots) subtraction of the expected noise, (a) from one act to the next, (b) from one loop to the next; the lines are regressions on the logarithm of the length. Spearman correlation between the mean distance and the length: -0.59 then -0.04 at the act grain, -0.30 then -0.02 at the loop grain. The noise made up 42% and 35% of the mean raw distance.

Table E1. The single model on the 1,770 pairs of 60 tables, corrected multi-scale d^2 , in units of the mean pair ($= 1$). Component = coefficient / 2 for the factors, coefficient $\times$ variance of the covariate in the corpus for the covariates (variance of the number of players 0.93; of the share of women divided by 0.25: 0.76). The coefficient of determination on the pairs is 0.073: the model explains the cell means, not the pairs one by one.

term	coefficient	standard error	95% interval	p (permutation)	component ($d^2 / 2$)	share of between-table variance
constant (table-specific)	0.354	0.127	[0.104, 0.603]		0.177	34% [10, 58]
different GMs	0.358	0.126	[0.112, 0.604]	0.0012 (0.0035 within system)	0.179	34% [11, 58]
different systems	0.236	0.063	[0.113, 0.358]	0.0001	0.118	23% [11, 34]
(Δ players) ²	0.036	0.013	[0.011, 0.060]	0.0089	0.033	6% [2, 10]
(Δ share of women / 0.25) ²	0.017	0.020	[-0.023, 0.057]	0.15	0.013	2% [-3, 8]

by permuting its labels between tables (9,999 times), the other terms remaining in the model, and counting the permutations in which the coefficient is at least as large as the observed one. For the GM, a second version permutes GMs within each system, which holds the system constant. The permuted units are the tables, exchangeable under the null hypothesis of no effect of the factor.

The intervals. Pairs are not independent (each table enters 59 pairs), and resampling must bear on the tables. Bootstrap with replacement creates pairs of a table with itself and duplicated pairs; we used subsampling without replacement (Politis, Romano, & Wolf, 1999): 500 draws of $m = 48$ of the $n = 60$ tables, refitting of the model on the remaining pairs, standard deviation of the estimates. A subsample that keeps 80% of the tables varies much less than the estimator itself, and the standard deviation must be rescaled by $\sqrt{m / (n - m)} = 2$, the jackknife factor for $d = n - m$ deletions (Shao & Wu, 1989). This scale was checked in three ways: by the delete-one jackknife (standard error of the GM's variance share 10.5 points against 11.2 by rescaled subsampling), by the standard deviation of the permutation distribution (13 points), and by a simulation of the additive model fitted on the real design with emulated measurement noise, in which the estimator of the shares is unbiased and its standard deviation is 10.6 points for the GM, 6.7 for the system, 3.5 for the players, 2.9 for the share of women and 11.9 for the table-specific share.

Appendix F. The power of the tests

Table F1. Minimum detectable effect at 80% power (in points of transition proportion, by interpolation on the power curve of Figure 4b), equivalent δ of the observed component, and share of rejections with no injected effect.

factor	form	minimum detectable effect at 80%	equivalent δ of the observed effect	rejections at $\delta = 0$
system	diffuse	2.0	2.5	4%
system	one transition	2.8	3.4	4%
GM	diffuse	3.6	4.0	5%
GM	one transition	4.4	> 5	4%
number of players	diffuse	1.1	0.9	0%
number of players	one transition	1.3	1.1	5%
share of women	diffuse	1.4	0.8	5%
share of women	one transition	1.9	1.0	6%

Power is measured by simulation, as recommended for designs with several random factors (Judd et al., 2017; DeBruine & Barr, 2021). For each factor, each form of effect and each size δ , one hundred replicates are simulated from the real count matrices. The factor's labels are permuted between tables, which erases the real effect while keeping the true noise and the true sample sizes. An effect is drawn according to these labels: for the system and

the GM, one value per category drawn from a normal distribution with standard deviation δ ; for the number of players, δ points per player above the mean; for the share of women, δ points per 25 points above the mean. It is injected into the matrices of each table by moving the corresponding share of probability from a source transition to a target transition, in two forms: “one transition” moves fiction $\rightarrow$ fiction towards fiction $\rightarrow$ game loop at the loop grain and player *asserts* $\rightarrow$ player *asserts* towards player *asserts* $\rightarrow$ GM call for a roll at the act grain; “diffuse” moves, for each row, δ points from its most frequent destination towards the game loop (towards the GM's game-level state that it most often joins, at the act grain). The distances are recomputed with the same noise correction and the same multi-scale distance, the single model is refitted and the factor's term tested by 199 permutations. Power is the share of replicates with $p \leq 0.05$; at $\delta = 0$ it is the type I error rate, which is between 0 and 6% depending on the test (at nominal level). The mean component obtained at each δ serves as calibration: the equivalent δ of the observed component is the effect size that would produce it on average. On the single fiction $\rightarrow$ game transition, the minimum detectable effects are somewhat larger than for a diffuse effect (Table F1). A significant effect obtained at low power, which is the GM's case, has a fair chance of being overestimated.

Crossed designs. The same simulation was run on hypothetical crossed designs in which each of M GMs runs each of S systems, one table per cell, each table being a real matrix of the corpus drawn at random with replacement, with a GM effect and a system effect of 2 points injected together in the diffuse form, and each factor tested with the other held fixed by stratified permutation. The power is 0.23 for the GM and 0.14 for the system on the real design (the 19 tables of the 7 GMs with several tables, 8 systems), 0.63 and 0.61 for a 5×5 crossing, 0.72 and 0.77 for 6×6 , and 0.92 and 0.93 for 8×8 : about seven GMs crossed with seven systems, some fifty tables, reach 80% on both sides (Section 10.6).

Appendix G. Sensitivity analyses and what tempers the GM's share of the variance

G.1. The cells of the design. Before any model, the 1,770 pairs of tables fall into four cells (Figure G1). Two tables with the same GM on the same system are at $d^2 = 0.43$ (3 pairs); two tables with the same GM on different systems at 0.67 (15 pairs); two tables on the same system run by different people at 0.79 (276 pairs); two unrelated tables at 1.04 (1,476 pairs). Changing the system adds 0.24, changing the GM 0.36, changing both 0.61, that is, the sum of the two effects with no visible interaction; and two tables with the same GM on different systems are somewhat closer than two tables on the same system run by different people, the standard error of the 15 pairs (about 0.09) covering most of the gap. These four cells confound everything that varies along with the GM and the system, which is what the model of Section 8 disentangles.

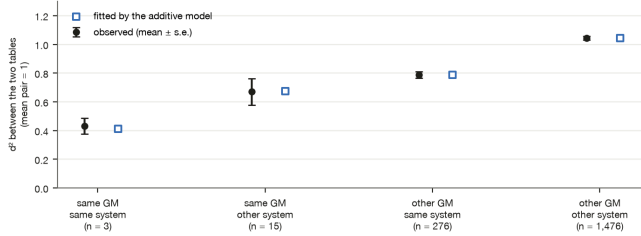


Table G1. Shares of between-table variance (%) by variant; in square brackets, the 95% interval when it was computed. The “different channel” term is 1 when the two tables are published by different channels; same GM implies same channel, and 25 pairs of tables from the same channel have different GMs (relative d^2 0.81 ± 0.38). The channel component is 0.104 ± 0.059 and the GM component beyond the channel 0.076 ± 0.076 (delete-one-channel jackknife). The pairs of tables with the same GM have relative d^2 : Gardien 0.25, 0.26, 0.33, 0.38, 0.43, 0.80; La Table Virtuelle 0.53, 0.53, 0.63; Wasitum 0.50, 0.68, 0.71; Jeux de Trolls 0.51, 0.53, 1.15; Abolas 0.69; bakadevil 0.73; La Réponse Dé 1.69. The noise of a pair is the term of Appendix D; pairs of short tables are the noisiest, and the correction removes the bias but not the variance.

variant	table-specific	GM	system	players	women
reference: 60 tables, acts and loops	34	34	23	6	2
act grain alone	44 [23, 66]	34 [11, 58]	10 [3, 18]	8 [3, 13]	3 [-3, 9]
loop grain alone	24 [-8, 57]	34 [4, 65]	35 [15, 54]	5 [-2, 11]	2 [-5, 9]
the 19 tables of the 7 GMs with several tables	24 [-4, 51]	32 [4, 59]	34 [3, 65]	9 [-1, 19]	2 [-5, 9]
without the GM of JDR du Gardien (56 tables)	43	25	22	8	2
without the GMs of Gardien and Jeux de Trolls (53)	44	26	19	8	2
without one of the five other GMs with several tables	32 to 34	34 to 37	22 to 24	5 to 7	2 to 3
with a “different channel” term	34	15 (+ channel 20)	23	6	2
regression weighted by 1 / pair noise	37	31	24	6	2
regression weighted by 1 / noise ²	41	24	27	6	2

Figure G1. The four cells of the design: corrected squared distance between two tables, in units of the mean pair, according to whether they share the GM and the system (black dots: observed mean and its standard error; blue squares: value fitted by the additive model, averaged over the pairs of the cell; n: number of pairs). Additivity can be read on the four points: $0.43 + 0.24$ (changing the system) + 0.36 (changing the GM) = 1.03 expected, 1.04 observed.

G.2. The variants. The eight pre-declared variants of the plan (uncorrected distances, tables of a thousand acts and more, classic systems only, systems with two tables and more, without the two-player table, share of women among players, each grain alone) do not change the conclusions of the model, with the exception of the variant without noise correction, in which all components are diluted by noise and the GM is no longer detectable (Appendix D). On the 19 tables of the 7 GMs with several tables, the system with the GM held fixed and the GM with the system held fixed are both detectable by stratified permutation ($p = 0.0004$ and 0.003) and of the same order.

G.3. What tempers the GM’s share of the variance. The GM component calls for three reservations. It rests on 18 pairs of tables and 7 GMs; 6 of these pairs come from the same GM, that of JDR du Gardien, and are the closest in the corpus (0.25 to 0.43 except one). Without this GM, the GM’s share of the variance falls to 25% and the table-specific share rises to 43%, whereas removing any other GM leaves it between 34 and 37% (Table G1). The GM’s and the table’s shares moreover compensate each other from one estimate to the next (correlation of -0.86 across subsamples): their sum, 68% (± 6 points), is much better established than their split. Finally, a GM with several tables is almost always the same channel, often the same crew, the same period and the same set-up. When a “different channel” term is added to the model, the GM component splits into a channel share (20%) and a GM share beyond the channel (15%, not distinguishable from zero); the 25 pairs of tables from the same channel run by different people are at 0.81, between unrelated tables (1.04) and same-GM tables (0.67). The corpus therefore does not separate the person who runs the game from the channel that publishes it. The table-specific share, finally, is an upper bound: an error of the chain that is random but constant within a table, a voice misattributed over a whole campaign or an editing habit, ends up in it rather than in the noise term of Appendix D.

Appendix H. The results without the out-of-game talk

The share of a table’s acts that are out of game (acts at the out-of-game level, or decoded in the out-of-game loop) ranges from 3% to 31% (median 10%); it is a trait of the table and of who runs it (over the seven GMs with several tables, intraclass correlation 0.62 , $F(6, 12) = 5.4$, $p = 0.007$), it grows with the number of players ($\rho = 0.46$, $p < 0.001$) and does not depend on the system (Kruskal-Wallis test, $p = 0.47$). The check of Section 8.4 reruns the whole analysis without any out-of-game act. In each session, every act at the out-of-game level and every act decoded in the out-of-game loop is removed (14,588 acts of 122,997, that is 11.9%); the act before and the act after become neighbours, as if the passage had been cut, and the rule that merges two consecutive identical acts by the same participant is reapplied (843 acts). The three out-of-game states leave the act grain, which goes down to 15 states, and the out-of-game loop leaves the loop grain, which goes down to three; the hidden Markov model is not refitted; the tables, the covariates, the distance, its correction, the tests and their seeds are identical.

Table H1. Shares of between-table variance (%) in the single model, with the out-of-game talk (Table E1) and without it; in square brackets, the 95% interval by rescaled subsampling of tables (Appendix E). p by permutation of the term’s labels between tables, with and without the out-of-game talk: GM 0.001 and 0.0008 (0.004 and 0.002 when GMs are permuted within systems), system 0.0001 and 0.0001, number of players 0.009 and 0.10, share of women 0.15 and 0.23.

component	with (reference)	without
GM	34 [11, 58]	42 [17, 66]
system	23 [11, 34]	30 [15, 46]
number of players	6 [2, 10]	4 [-1, 8]
share of women	2 [-3, 8]	2 [-3, 7]
table-specific	34 [10, 58]	22 [-1, 46]

The shares of the GM and of the system increase and the table-specific share decreases: the out-of-game talk is mostly variance specific to each table, and removing it cleans the distances rather than creating them. The gap between the GM and the system does not change (11 to 12 points, with the same uncertainty), nor does the power of the tests (minimum detectable effect in the diffuse form: system 2.0 points, GM 3.7, number of players 1.2, share of women 1.6, against 2.0, 3.6, 1.1 and 1.4 with the out-of-game

talk). The share of women remains indistinguishable from zero. The number of players keeps its share at the act grain (8% → 9%, $p = 0.0001$), which is the arithmetic of turn alternation, and loses its share at the loop grain (5% [-2, 11] → -1% [-11, 9]): more players means more out-of-game talk, and the entries into and exits from that loop carried the effect; at the two grains together, its share goes from 6% to 4% and is no longer distinguishable from zero. What weakens are the tests with the other factor held on the 19 tables of the seven GMs with several tables (Appendix G): the system with the GM held goes to $p = 0.08$ and the GM with the system held to $p = 0.12$, against 0.0004 and 0.003 with the out-of-game talk; these tests have almost no power and lose 12% of their material. The check answers the question of whether the results hold without the out-of-game talk, not what an editing that also cut into the game, a rules discussion that drags on or a scene that fell flat, would do, which nothing allows us to simulate.

Appendix I. The model arrow by arrow

The method. For each arrow of Figure 3 (a transition within a loop, a passage from one loop to another, an entry into or an exit from a loop through a given act), the proportion is recomputed for each of the 60 tables, smoothed towards the corpus value with the same weight as in Appendix D. The squared difference between two tables, minus the sampling noise of the two proportions (the expectation of $(\tilde{p}_i - \tilde{p}_j)^2$ is $(p_i - p_j)^2 + v_i + v_j$, v being the sampling variance of a smoothed proportion), is regressed on the same four terms as in Appendix E: different GM, different system, squared difference in the number of players, squared difference in the share of women. The coefficients are the components of the between-table variance of that one proportion, and the p -value of each term comes from shuffling its labels between tables, as in Section 8.2 (1,999 permutations, all arrows at once). Only the sign of the players' effect is lost by this regression, which estimates a squared slope; it is recovered as the slope of the proportion on the number of players with the system held fixed. In Figure 5, an arrow is coloured by the factor with the largest component among those established on it at $p \leq 0.05$; the number on it is the square root of that component, in points; its width is proportional to the component.

Two cautions. First, the colours say what is established on an arrow, not what is largest: in size, the GM has the largest component on 37 of the 52 arrows, but its test on a single arrow has little power, since it rests on the 18 pairs of tables with the same GM (Appendix G). On a typical arrow, the smallest effect detected four times out of five is about 9 points for the GM, 6 for the system and 4 for the number of players; a 5-point players' effect is therefore coloured where a 7-point GM effect is not. Averaged over the 52 arrows, the components give an apportioning close to that of Figure 4 (GM 45%, system 14%, number of players 11%, share of women 5%, table-specific 26%). Second, 52 arrows tested against three factors at the 5% level would yield about eight coloured arrows by chance alone; the players' arrows are well beyond that (7 of the 10 at $p \leq 0.01$), the GM's less so (3 of 8) and the system's least (1 of 6), and no single arrow should be read as established on its own.

The coloured arrows. The number of players is established on 10 arrows: a player's fiction act repeated by another player (+5 points), the in-character line repeated (+7), a player's question followed by another player's assertion (+4), the passage from the out-of-game loop to the fiction (+4), the entry into the character loop through a player's line (+5); a player's assertion or question in the fiction followed by the GM's assertion (-6 and -6), the entry into and the exit from the character loop through the GM's line (-5 and -6), the exit from the game loop towards the fiction (-3). Over the whole corpus, the share of player acts followed by a GM act falls by 6.9 points per additional player

($r = -0.57$ between tables). The system is established on 6 arrows: the opening of the game loop by the GM's call for a roll (9 points; that entry makes 45% of the entries into the game loop on *The Haunting* and 49% on *Candela Obscura*, against 21% on *Lost Mine of Phandelver* and 14% on *Root*), the passage from the fiction to the game loop (8; *The Haunting* enters the game loop from 2.6% of its fiction acts, *Chariot of the Gods* from 4.3%, the "apocalypse" games from 4.4%, *Death House* from 6.7% and *Phandelver* from 7.4%) and from the out-of-game loop to the game loop (6), a player's request at the game level followed by the GM's assertion (7), and, outside the game loop, the line of a player's character answered by the GM's character (8). The GM is established on 8 arrows: the repetitions out of game (9 and 10), and the exchanges within the game loop, a player's assertion at the game level followed by the GM's (14), repeated by another player (10), or closing a passage through the game loop (12). The full table of the 52 arrows, with the four components and their p -values, is written by the tool listed in Appendix K.

Appendix J. The passages through the game loop and the resolution styles

J.1. Three measures on the passages through the game loop. A passage through the game loop is a maximal run of consecutive acts decoded in that loop. The *take-over rate* is, among the transitions within the game loop that leave a player's statement at the game level (the merged state "P asserts + game"), the share whose next act is the GM's assertion at the game level; it is the transition of Figure 5 on which the GM is most clearly established. The *narration length* is the median number of words of the GM's first assertion at the fiction level among the three acts that follow the end of a passage. The *player-authored return* is the share of passages for which that first assertion at the fiction level is a player's. Over the 58 tables with at least 15 player statements at the game level, the take-over rate ranges from 17% to 93%, the narration length from 10 to 82 words and the player-authored return from 12% to 52%; the take-over rate is unrelated to the narration length (Spearman's $\rho = 0.05$) and weakly to the player-authored return ($\rho = -0.11$). For the seven GMs with several tables, take-over rates, narration lengths and player-authored returns are, table by table: Les JDR du Gardien 76, 78, 70%; 37, 47, 33 words; 22, 24, 14%. Abolas 68, 63%; 17, 26 words; 29, 32%. La Table Virtuelle 63, 71, 54%; 10, 36, 39 words; 41, 45, 31%. Jeux de Trolls 56, 33, 53%; 26, 32, 29 words; 35, 31, 29%. Wasitum 30, 53, 48%; 55, 42, 43 words; 34, 14, 21%. La Réponse Dé 25, 53%; 28, 25 words; 39, 17%. bakadevil 30, 33%; 16, 25 words; 47, 38%. A reading of passages drawn at random from tables at the two ends of the take-over rate, on the same scenario, is what suggested the styles of Section 9.3: a GM who narrates every result ("one success, so you shoot the thing flying from the android's hands, you see a spurt of blood, a strangely yellow blood"; a point of stress gained by shooting someone becomes "you are shocked, in fact, by shooting someone"), a GM who records it ("those who failed their roll lost as many units of consumables as they rolled ones; at zero, you have no more air"), a GM who hands it back ("you manage to break free, you take two steps back; how do you react?"). The reading is not a measurement, and the annotation of Appendix B does not distinguish a narration that translates a result from any other narration; the codebook below was written to code it.

J.2. The resolution styles: unit and codebook. A *resolution passage* is a maximal run of at least two consecutive acts decoded in the game loop in which at least one game-level act, by anyone, contains a number or a result word (39% of the candidates have that marker only in the GM's mouth, which rules out filtering on players' statements); the corpus contains 1,984 of them (median 26 per table), and the coder reads each with two acts of context before and six after. The label comes from ordered questions. Is a result actually stated in the window? If not, the passage is *no reso-*

lution (19% of the sample: equipment, bookkeeping, rules explained, rolls called but never made). Does anyone state a fact about the world that follows from the last result before the conversation moves on? If not, *accountant*. Who states it first, and does the GM let it stand? A player's statement left standing is *delegated*. Does the GM's statement contain anything the outcome does not imply, a detail, a reaction, a consequence? If so, *narrator*; if not, *laconic*. Boundary rules settle the frequent cases: the rulebook's phrase read aloud is *accountant* even when it speaks of the character; a player who restates the verdict of the dice in fiction words ("it slips", "I can't reach it") is not thereby the author of the outcome; on perception rolls, a minimal piece of information is *laconic* and a developed one *narrator*. Side fields record the type of roll (attack, damage, action, perception, suffered, the GM's own dice, other), who announces the result, whether the stakes were stated before the roll, whether the GM invited the player to narrate, whether the outcome adds a consequence beyond success or failure, and the coder's confidence.

Table J1. The four resolution styles, defined by the first statement about the world that follows the last result of a passage through the game loop (Section 9.3); examples translated from the French. Shares are among the 162 of the 199 coded passages that carry a result; the other 37 carry none (equipment, bookkeeping, rolls called but never made).

style	the first statement about the world after the last result...	example	share
narrator	is the GM's and contains at least one element that the outcome does not imply	GM: "There is something at the base of your neck that dislocates."	62%
accountant	does not come: the result stays a game state, points lost or the rulebook's phrase read aloud, and the fiction resumes elsewhere	GM: "Those who failed their roll lost as many units of consumables as they rolled ones."	16%
laconic	is the GM's and restates the outcome in a single bare proposition	GM: "You activate the thrusters and free the ship."	14%
delegated	is a player's, who says what the success becomes, and the GM lets it stand	P: "Fourteen damage. I push off you, I jump, I take my run-up with my axe and land a huge blow on its flank." GM: "You cut off one of its wings? Yes."	7%

J.3. Design and coding. The 25 tables are the 19 tables of the 7 GMs with several tables, whose crossing graph is connected, plus 6 tables chosen at the extremes of the number of players (2, 3 and 6) in systems already present; 8 passages per table (one table has 7), half of them drawn among the passages whose acts carry attack vocabulary (30% of the candidates; 40 to 50% in the *Dungeons & Dragons* scenarios against about 20% elsewhere), so

that a contrast between systems is not a contrast between mixes of rolls; 199 passages, 55,000 words. A simulation on this design shows that the net permutation tests detect a between-GM spread of one logit about one time in five: the sample calibrates the codebook, it does not measure the shares. The coder labelled 60% of the passages without hesitation; the hesitations concentrate on the *laconic/narrator* boundary, whose threshold (one element the outcome does not imply) is low.

J.4. Results. Of the 162 passages with a result: narrator 62%, accountant 16%, *laconic* 14%, *delegated* 7%. By GM, the narrator share is 100% at the four tables of *Les JDR du Gardien* (two systems), 86 to 100% at the three tables of *Wasitum* (three systems, two of them "apocalypse" games), 62 to 71% at the three tables of *Jeux de Trolls*, 71 and 88% at the two tables of *La Table Virtuelle* with at least seven resolved passages; *Abolas* is the exception, 80% on *Chariot of the Gods* against 20% on *Death House*, five passages each. The tests fit one logistic model per outcome at the passage grain with the type of roll held, measure the deviance drop when the factor is added, and draw the p-value from 999 permutations of the factor's labels between tables: the GM within the system gives $p = 0.002$ (narrator), 0.001 (accountant) and 0.011 (a score of fiction produced: 0 accountant, 1 *laconic*, 2 narrator or *delegated*); the system within the GM gives $p \geq 0.85$ on those same outcomes, the number of players within the system $p \geq 0.56$. The single model of Appendix E, applied to the narrator share per table with the squared difference corrected for binomial sampling noise, puts the GM's component at 75% of the between-table variance with an interval of $[-49, 198]$ and $p = 0.007$ when permuted within the system: at 25 tables and six resolved passages per table, only the permutation verdict is informative. In the side fields, the stakes are stated before the roll in 55% of the *Blades in the Dark* passages (its position-and-effect procedure) against at most 9% elsewhere, and a consequence is added beyond success or failure in 71 to 86% of the "apocalypse" passages against 35 to 59% of the classic ones; the one GM with enough passages on both sides goes from 0% of added consequences on *Chariot of the Gods* to 57% on *Monster of the Week*, in the direction the moves prescribe, but another adds consequences at all their tables (71 to 100%), and neither contrast survives the permutation of systems within GMs ($p \geq 0.20$). The 12 *delegated* passages are 20% of the attack rolls and at most 5% of any other type; one table, a duo, organises the delegation by invitation ("after you", "tell me how it goes", 38% of its passages). The next step is the annotation of the 1,984 candidate passages with a widened window and a graded level of fiction produced (none, state, terse, described) in place of the threshold, and the comparison of several annotators on these 199 passages.

Appendix K. The tool chain

Table K1. Everything is rerun from the corpus folder; commands and conventions are in the study's README and in the pipeline note.

step	input	tool	output
audio	published video	yt-dlp, ffmpeg	16 kHz sound
transcription	sound	Whisper large-v3-turbo (MLX implementation), completion pass	time-stamped words
voices	sound	pyannote 3, ERes2NetV2 embeddings, enrolment by ear	attributed segments (GM, P1, P2...)
gender	segments	Praat (Parselmouth), wav2vec 2.0 age-gender, Gaussian mixture on embeddings	gender per voice, share of women per table
turns	transcription, voices	merge_turns.py, gm_speech_share.py	speaking turns, GM's share of speech (Section 3.4, Appendix A)
acts	turns, codebook and example	gpt-5.4-mini by slices of one hundred turns	one or more codes per turn
sequences, merged states	acts	act_sequences.py, merge_states.py	18 states, repeated acts merged (Table C1)
graph, matrix	sequences	figure_transition_graph.py --sans-legende --bulles-egales --langue en, figure_transition_matrix.py --langue en	Figure 2, Figure C1
loops	sequences	loops_hmm.py (hidden Markov model, BIC), loop_structure.py, figure_nested_loops.py --sans-legende --sous-titre court --langue en	four loops, decoding, Figure 3
tables, corrected distances	decoding, covariates	four_questions.py --etape matrices,distances	matrices per table, distances
single model	corrected distances	distance_regression.py, distance_regression_power.py, article_figures.py --langue en --journal	Table E1, Table F1, Figure 4, Figures D1 and G1
checks	corrected distances	check_central_result.py, crossed_design_power.py (archive)	Appendices E, F and G
without out-of-game talk	decoding, covariates	four_questions.py --variante sans-hors-jeu, distance_regression.py --variante sans-hors-jeu, compare_variants.py	Section 8.4, Appendix H
which factor moves each transition	decoding, covariates	arrow_effects.py, figure_nested_loops.py --effets --nombres points --epaisseur variance --niveaux 1 --legende-effets couleurs, rules_excerpts.py --mesures	Section 9, Figure 5, Appendices I and J
resolution styles	decoded passages	resolution_passages.py, resolution_sample_analysis.py, resolution_sample_shares.py	Section 9.3, Appendix J, Table J1
chain diagram		figure_pipeline_diagram.py --langue en --colonnes 3	Figure 1