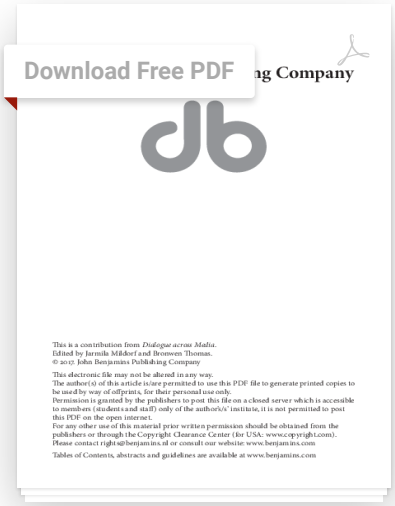

Dialogue in video games

Dialogue Studies, 2017

Sebastian Domsch

movement of a joystick or a mouse click is, after all, an act of communication to which the game reacts – but the impression that many games create of actually talking to a sentient being can only ever be an illusion (we will look at the exception of player-to-player conversation later).[1] Conversations between human beings, even the most unimaginative ones, have a level of emergence that no game system could yet hope to match. Instead, what games offer are simulacra of conversations that are subject, to a varying degree, to modification by the player. Many of these, as we will see later, are not even very convincing, and some are downright ridiculous and artificial. And yet, our yearning to comprehend our actions in anthropomorphic terms – even or particularly when they are only simulated – makes the use of language as a means of communication enduringly appealing. Text-based messages have been present in games from the start, and as soon as the development of game technology allowed them to include voices, they did.

One phenomenon, or recurrent structural feature, that encapsulates the paradoxical nature of interactive dialogue in video games, is the so-called "silent protagonist." This refers to a player character who, throughout the entire game, has little or no dialogue, whether in written or in spoken form. While this is not so significant in a *Super Mario* game or in *Space Invaders* (where nobody would expect the avatar to speak), it is much more noticeable in games that create a complex and detailed fictional world. The reasons for the predominance of silent protagonists are partly technical, as especially earlier games did not have the means to give the player character a voice. But the most important reason is that it eases identification and therefore immersion, as C. J. Miozzi (2012) explains: "players seeking escapism wish to take on the role of a story's hero, and a voice other than their own jars them from this fantasy. By leaving the protagonist a clean slate, the player can fill in the blanks as he sees fit." However, this feature has come under increased criticism lately. One point of criticism is that it is failing to fulfil its own purpose: "In RPGs in which interaction is paramount, that lack of characterization actually serves to break immersion due to the jarring disparity between voiced NPCs[2] and the silent protagonist" (Miozzi 2012). And furthermore, silent protagonists are seen as a weak solution for the implementation of narrative into video games: "Silent protagonists in video games have become the voice-over narration in movies – a device used as a crutch to compensate for bad storytelling" (Miozzi 2012).[3] In order to understand

1. On this point of human-machine communication on the one hand and communication between players on the other, see also the contributions by Piwek and Mäyrä in this volume.

2. An NPC is a non-player-character, i.e. every agent within the gameworld that is not the player's character or avatar.

3. On the silent protagonist, cf. also Vanden Bossche (2008) and Redsell (2011).

the different forms of dialogue that can occur in video games, it is important to first look at the different levels of communication that can be involved.

two separate levels of communication, one that is understood to happen within the fictional world that the game presents (which we might call diegetic communication), and one that is directed at the player (which would be ludic communication). To complicate things further, these levels can even be present simultaneously, or even be merged in a single utterance that is understood to refer both to the fictional gameworld and to the actual gameplay.

Ludic communication addresses the player as player. It acknowledges the fact that the player is playing, and its content is usually about the activity of playing. A basic example would be the message "Game Over. Enter Coin to Continue," common to arcade machines. As this example shows, such communication is often not very complex, and the player's options for answering rarely involve language (beyond a simple "yes" or "no"). The topics for this communication are always references to the game's rule structure. In a non-active game medium like chess or *Monopoly* (where the game system cannot do anything that is not a direct reaction to a player's input), such communication about rules is usually distinct from the game object proper. But since games are an active medium, they can also communicate rules on their own, whenever it seems appropriate. When the player makes a wrong move, the game system will not only not allow the move, but add a message saying "You cannot do that!" – even if it is just through a sound. Similarly, it can tell the player about options that are available at a specific moment, without having to present all of the options at the outset. It is this feature which makes the often incredibly complex rule structures of video games possible: video games can "talk" to their players, teaching them how to play the game *as* they play the game.

Besides game-to-player communication, many recent multiplayer games also allow their players to communicate with each other. This happens usually through chat channels and sometimes through voice-over IP technology, so that players can actually hear each other talk. As a tool that enables free communication between players, this communication usually lacks any necessary connection to the game. Players are free to talk about the weather or recent political developments. Of course, theoretically players can decide to move their dialogues into the diegetic sphere of the game, that is, talk "in character," as if they were their own avatars. But not only is this not what happens in the majority of cases, it would then also constitute a clear example of role-playing, which is covered by Frans Mäyrä in this collection.

More relevant to my purposes are examples that attempt to tie player-to-player communication in a more direct way to the game experience. One such example would be the game *Demon's Souls* (2009). Structurally, this is a single-player game, in that every player inhabits their own game space, but it does allow a lower-level multi-player element in so far as players can leave messages on the floor of their game space (usually warnings about traps) that can be seen by other players in their own game space.

Another interesting example is the game *Journey* (2012). In this game, the player controls a robed figure in a large and empty desert landscape moving towards a mountain that is visible in the background. On this journey, the player can encounter other players (only one in any given area at the same time), who will appear as similar figures, but without any name or identifying features. The two players can then journey together and assist each other. The only way that players have to communicate with each other is through musical sound, which also becomes visible as a sign on the screen. In *Journey*, it is the reduction of player-to-player communication and the exclusion of language that ties this communication to the game's fictional world. The player is more or less forced to understand his/her communication with the other player – who can only be known through a nameless avatar, and not, as in most other games, through a nickname that the player has given herself – as a dialogue between the two characters travelling through the game's desert landscape.

Ludic conversation, as we have seen, is directed at the player and refers to the activity of playing. More recently, games that emphasise immersion in the game's fictional world have tended to hide the ludic nature of this type of communication by putting it into the mouth of a figure that also exists on the diegetic level

be omitted. The dialogue happens completely within the game's diegesis and is understood as being part of it. This accounts not only for the majority of all cases of dialogue in video games, but also for the structurally most interesting ones. As we have seen, direct ludic communication tends to be simple and one-sided (basically the game system giving commands or prohibitions), but diegetic communication cannot only use a broad range of expressive modes, it can also create deep, extended, and branching dialogues. This leads us to the heart of dialogue in video games, and I will mainly concentrate on the forms and functions of diegetic dialogue in the following.

In addition to direct communication where the player character talks to an NPC, one feature common to recent games that emphasise the creation of what marketing departments like to call a "living, breathing world" is that NPCs will

often speak or talk to each other without directly addressing the player character. The effect is that of a conversation overheard randomly, or so it seems. The most ambitious form of this kind of dialogue will adapt the overheard conversation to the course of events or the gameplay up to that point. Ideally, the player will form the impression that, by lucky coincidence, she has stumbled upon two people who are talking *about her* without knowing that she can hear them. This is a common element to stealth games like *Assassin's Creed*, where the emphasis is on the player entering or exiting locations without being noticed by NPCs (in contrast to combat games where the player would have to shoot every NPC that is in her way). This creates situations where the player character is silently sneaking past unaware NPCs, giving them time to talk with each other, and the subject of these overheard conversations will often be the player character.

But while theoretically enhancing immersion in the storyworld, this semblance of random conversation can easily tip over into incredulity, by either becoming repetitive (the same line of dialogue is spoken "randomly" again and again) or ridiculous, or both. Such was the case in the game *The Elder Scrolls V: Skyrim* (2011), where the player character would frequently meet with (different) characters who would tell him without prompting: "I used to be an adventurer like you, then I took an arrow in the knee." The latter part of this phrase quickly took off as a meme in the form of "I used to X, but then I took an arrow in the knee" and numerous image macros and video parodies were created.[4]

While ludic communication has the function of giving the player feedback about her actions and of informing her about the game rules, diegetic dialogue is all about the representation of a convincing and immersive fictional world in which the player's actions have significance. The main functions are what could be called exposition, description, and progression and are related to the narrative past, present, and future.

Forms of dialogue in video games

Technically, dialogue is represented in video games in two ways, as either written or spoken language. Historically, written language presented statically on the screen (often in the form of comic-like speech balloons) was predominant until the arrival of the CD-ROM as a storage device enabled game platforms to hold larger quantities of high quality audio files. Early examples show the difficulties of implementing spoken language in video games, like *Berzerk* (1980), notably the

4. Cf. http://knowyourmeme.com/memes/i-took-an-arrow-in-the-knee.

first game to use speech synthesis. It allegedly cost $1000 for each word to process the LPC coding (cf. "Berzerk"). Still, written text-based language continues to be used today, since it is so much easier to produce than spoken language, which needs the text and actors, recording technology, storage, etc. Genres that rely heavily on dialogue but are not AAA productions with a budget worthy of Hollywood, such as visual novels, are especially dependent on text.

Before the advent of complex and convincing graphics, some of the earliest video games relied completely on text, creating a genre that was appropriately called "text adventure." But the main role of the text was to be a substitute for the visual dimension, and therefore text was overwhelmingly descriptive. Also, players would usually not engage directly in conversation with NPCs, but would only communicate with the game system. It was only when graphics freed the word from the burdens of representation that games could focus more exclusively on dialogue.

Even further along the line, games started to substitute written with spoken text. Dialogue acted out by performers can of course add important layers of meaning to the dialogue's words (e.g., through paralinguistic information such as the type or tone of voice, or emphasis), although it also entails the danger of reducing credibility through bad performance. While this was quite an issue in the earlier stages of video games' development, today voice acting for video games has been professionalised and its specific challenges are widely acknowledged. These challenges lie mainly in the non-linear and strongly decontextualised nature of the production process, where actors have to perform vast amounts of individual lines of dialogue which do not form a single development, but provide a range of sometimes widely differing options.

While the voices of actors have broadened the appeal and (in the good cases) strengthened the immersive qualities of dialogue in video games, the speaking faces that the player would see on the screen have still mostly been rather wooden animations or cartoonish abstractions. One game that marked the possible shape of things to come is *L.A. Noire* (2011), which used sophisticated motion capturing to transpose not only the actors' words but also their facial movements into the game. And this feature is not only used to enhance the game's production value, but is an integral part of the gameplay, since the player takes on the role of a police investigator who frequently questions suspects and therefore often has to judge whether her dialogue partners are lying or saying the truth. This assessment can often not be made with reference to the content of the words alone but has to take into account all the non-verbal signs that a real-life conversation adds to these words. Players have to compare what a suspect says with the way he says it and even with his facial expression – does he keep eye contact and an open face, or does he look down and bite his lips nervously afterwards?

Not all dialogue in video games contains choices for the player, even though they might be interactive. Especially in the frequent cases where the player does not even have a choice as to whether to engage in dialogue at all, and then maybe not even answers to choose from, the "interactivity" (where the player has to press a button for the dialogue to continue) is not optional, so this type is only as interactive as a DVD insofar as I can choose to start or not. It is merely a different way of presenting the dialogue, where every line appears only after a click. An even less engaged version would have a cut-scene with dialogue play after a specific player input. This is where dialogue in video games becomes indistinguishable from that in film or a book (with the clicks that move the dialogue forward akin to a turning of the page). I will therefore focus on those occurrences of dialogue that do enable player's choice in one way or another.

One major structural difference is whether the game system uses language merely as invariable content to fill what is in essence a branching multiple-choice structure, or actually tries to process the player's language input. Of course, true mastery of language comprehension would be a noteworthy feat for an accomplished artificial intelligence, and it has not yet been successfully realised. But

and transformed it to something the game could understand. Usually, words with the same meaning are turned into the same word (e.g., "take" and "get") and certain filler words are dropped (e.g., articles, or the "at" in "look at rock"). Thus, the player had a seemingly unlimited number of options for input, and the game, depending on how well the parser was created, could react meaningfully to all of them (see also Piwek, this volume).

The first time a text parser was used in a video game was in 1976 in the text-only genre-defining *Colossal Cave Adventure*, designed by programmer and caving enthusiast Will Crowther. But it was only with the advent of the game *Zork* (1977) one year later that the parser found a more advanced form. Whereas *Colossal Cave Adventure* would only recognise simple verb-noun compounds, *Zork* could even recognise some prepositions and conjunctions. To understand the excitement of such technical advances, one might cite a review of the game from 1981:

> I was eager to test Zork's biggest selling point, intelligent input (ie: its ability to accept free-form instructions). I typed "OPEN THE BAG AND GET THE LUNCH," in reference to a brown paper sack inside the house. The computer complied. There was water and food, so I typed "EAT THE LUNCH AND DRINK THE WATER," to which the computer responded with gratitude for satisfying its hunger and thirst. I was hooked. (Liddil 1981, 262)

Text parsers were anything but perfect. The main problem was that the designer had to anticipate the possible inputs of players and to create puzzles in such a way that players would be able to find the right words for the solution. In the worst case, this led to what has been called a game of "guess the verb," where the computer would only accept a specific command out of a range of possible synonyms. The most notorious example of this was found in *Leisure Suit Larry Goes Looking for Love (in Several Wrong Places)* (1988), where the player character had to put a bag in a bottle. But since in this case the parser reacted to input like "put bag in bottle" by interpreting "bag" as a verb, only the (usually unnecessary) inclusion of "the" before "bag" would yield the desired results, driving a whole generation of pre-internet players mad in the process.

Text parsing fell out of fashion as the mouse became established as an input tool and players could select objects for interaction by merely pointing and clicking. Ever since, games have relied on either direct (non-verbal) interaction through mouse or joystick or multiple choice options (thus, in essence, another point-and-click interaction) for verbal communication. It is only in most recent years that games have again started to explore the possibilities of linguistic input and language comprehension through the use of chatterbots or chatbots. These are computer programs that attempt to simulate chat communication, usually trying to pass the Turing Test, so that the dialogue partner will not recognise that she is talking to a computer.

One example that shows the potential of such programs as well as their practical difficulties is the mobile game *Boyfriend Maker* (2013). This is a variation of the dating sim genre, in that there is only one potential boyfriend, and instead of a branching but pre-scripted structure, the game parses the verbal input of the player and tries to react in a convincing way. To improve its capabilities, the game even used vocabulary and knowledge that it acquired in a chat with one player for chats with other players. While this apparently made some of the answers surprisingly convincing and lifelike, it also led to strange and sometimes disturbing incoherences. The game was designed to encompass the whole communicative range from flirting to sexting, and players became increasingly annoyed when the game would suddenly mix registers. Thus, after initial success, Apple deleted the game from its store (Hawgood 2013).

If dialogue is not presented as part of a cut-scene, it often makes more than one dialogue option available to the player, thus creating what is commonly referred to as a dialogue tree. The choice is generally restricted to the dialogue options of the player, and the non-player character will in turn react differently to different statements or questions. The dialogue choices are presented in written form and are arranged on the screen simultaneously, while the action pauses until one option

In the case of multiple options one major difference is whether the player knows exactly what words the player character will say for each option that is chosen or not. The first case is still the default one, although it does have practical disadvantages, such as a slowing down of the gameplay. In general, the more strategically oriented a game is (in contrast to action-based gameplay), the less of a problem a complete rendering of the dialogue will seem to be, or will even be regarded as indispensable. This is especially true if the strategic element includes, or is even centred on, how the player character interacts verbally with her environment. Thus, in a dating sim, where everything revolves around the emotional significance of dialogue, it is very important for players to know exactly the wording of each dialogue choice. In this case, judging the potential effect of a sentence ("If I tease, will he think I'm flirtatious? Or if I don't tease, will he think I'm boring?") is one of the appeals of the gameplay. In other cases, players might only want to choose an effect (dismiss/accept proposal; offer help) without being interested in the actual wording.

Since having to read all of the different dialogue options can disrupt the flow of the gameplay, many games have tried to find ways to abstract the different options and to present players with reduced information. Most games do that by focusing on either the mood that a choice expresses (angry/pacifying/enamoured), or the way a choice is expressive of a personality (jocular/gruff/upright/mischievous). The science fiction-themed role-playing game series *Mass Effect* developed such a system of indirect choices, which are represented as paraphrases of what the player character will actually say. Thus, the choice that the player can read might be "Don't try to study me," while the actual spoken line is "I'm not some artifact you can take back to your lab, doctor." In addition, dialogue options are arranged not merely as a list, but are placed on a wheel, with the relative position on the wheel also indicating the spirit or mood of the answer:

> The left side of the wheel is normally reserved for options that will continue the conversation in depth, including Charm and Intimidate options (see below); occasionally an "Investigate" option is given in the middle, which allows Shepard to ask about multiple topics. The right side of the wheel is divided into three sections, and tends to move the conversation towards quicker completion. Paragon responses, generally more selfless or cooperative, are on the top segment, the middle segment presents a more neutral option, and Renegade responses, often more aggressive and hostile, are on the bottom segment. ("Dialogue")

There are, of course, many variations. In the independent game *Will Fight For Food: Super Actual Sellout: Game of the Hour* (2015), players can choose between different options for "body language" (e.g., aggressive/defensive/relaxed/excited), "opinion" (e.g., agree/indifferent/disagree), and "tone" (e.g., deadpan/sincere/insult/sceptical), and the game will then choose a dialogue option according to the combination.