CONFLICTING MUSIC AND DIALOGUE IN POPULAR TV SHOW SCENES THROUGH THE LENS OF MULTILEVEL GROUNDED SEMANTICS [a]

Anja Tošić*, Mihailo Antović

University of Niš, Faculty of Philosophy, Niš, Serbia

ORCID iDs: Anja Tošić   https://orcid.org/0000-0001-6552-185X
            Mihailo Antović   https://orcid.org/0000-0002-5308-6747

Abstract

This paper aims at discussing the process of multimodal meaning construction called *multilevel grounding,* according to which meaning emerges hierarchically, through six recursive levels of constraint: *formal*, *cross-modal*, *affective*, *conceptual, culturally rich,* and *individual*. Given that this theory has found application in the cognitive science of meaning generation in music, language, and visual art, this study aims to investigate more thoroughly the nature of meaning construction when linguistic and musical data are cointegrated. To that end, we analyse 5 TV show scenes in which the dialogue and the tones heard in the background seem to be sending out contradictory signals on one/some of the six grounding levels this theory presupposes. By doing so, this study attempts to come one step closer to learning in what way language and music, even when they appear to communicate a conflicting message, still work together to create novel meaning. The results of the present study indicate that such linguo-musical mismatches occur on each of the five levels of interpretation, with the sixth level being excluded from the analysis on purpose due to its inherently subjective nature. The same phenomenon has been observed in each of the five analysed scenes, and we believe that its use was both intentional and necessary on the part of the shows' creators to communicate the intended message in its entirety to the keenest of viewers. Namely, we argue that it is exactly these contrasts on each level of constraint that are responsible for the complex understanding of the scenes in question by introducing entirely new elements into the narrative, such as plot twists, which further leads to often humorous, but always non-composite, interpretations.

Key words: meaning construction, multilevel grounding, music, language, conflict.

МУЗИКА И ДИЈАЛОГ СУКОБЉЕНИ У СЦЕНАМА ИЗ ПОПУЛАРНИХ ТВ СЕРИЈА КРОЗ ПРИЗМУ СЕМАНТИКЕ ВИШЕСЛОЈНОГ УТЕМЕЉЕЊА

Апстракт

Овај рад се бави процесом конструисања мултимодалног значења под називом *вишеслојно утемељење*, на основу кога значење настаје хијерархијски, кроз шест рекурзивних нивоа интерпретације, названих *формални, кросмодални, афективни, појмовни, културолошки* и *индивидуални*. Имајући у виду да је ова теорија нашла примену у делу когнитивне науке који се бави стварањем значења у музици, језику и визуелној уметности, ово истраживање има за циљ да дубље истражи природу изградње значења приликом комбиновања језичких и музичких информација. У том смислу, анализирамо 5 сцена из ТВ серија у којима дијалози и позадински тонови наизглед шаљу контрадикторне сигнале на једном или више нивоа утемељења које ова теорија претпоставља. На тај начин, овај рад настоји да приђе корак ближе разумевању начина на који језик и музика, чак и када се чини да шаљу контрадикторну поруку, и даље делују заједно како би створили темељ за ново, креативније тумачење. Резултати до којих смо дошли сугеришу да таква језичко-музичка неслагања постоје на сваком од пет нивоа интерпретације, при чему је шести ниво намерно изостављен из анализе због своје суштински субјективне природе. Овај смо феномен препознали у свакој од пет анализираних сцена и верујемо да су га аутори серија употребили као нужне како би се најпажљивијим гледаоцима у потпуности пренела намеравана порука. Дакле, у овом раду тврдимо да управо ти контрасти, присутни на сваком нивоу утемељења, пружају мотивацију за сложено разумевање ових сцена тако што у оквиру комплексног наратива најављују потпуно нове елементе попут неочекиваног развоја догађаја. Ово даље води до често духовитог, а свакако некомпозитног, тумачења сцене.

Кључне речи: изградња значења, вишеслојно утемељење, музика, језик, конфликт.

INTRODUCTION

Following decades of linguistic underdeterminacy present in semantic/pragmatic thought that recognised the value of context, background knowledge, and other extralinguistic factors, suggesting that linguistic information alone is not enough to determine the meaning of what is being said, we seem to have relatively recently witnessed a leap toward influential theories dealing with more than literal denotations of words. In particular, with Lakoff's (2006) idea that metaphor comprises a huge portion of everyday language and Johnson's (2005) claim that many human concepts are based on preconceptual 'image schemas', interest suddenly grew in attempting to analyse contextual aspects of meaning construction. More modern semantic theories are thus concerned with phenomena not necessarily bound to language alone but related to general cognitive processes and ultimately manifest in visual and auditory domains as well as language. George Lakoff's (2006) conceptual metaphor theory and Fau-

connier and Turner's (2002) theory of conceptual blending have thus proposed that concepts exist outside of language as well, and that, to analyse them, scholars need to resort to cross-field parallelisms between linguistic and non-linguistic information. For instance, the latter paradigm can be used to explain how novel meaning emerges from the conceptual integration of distinct textual and musical structures (Antović & Tasić, 2011).

While these theories analyse purely conceptual language phenomena, or phenomena delineated by narrow linguistic context, well enough, they still remain somewhat vague in that they fail to take into account the indisputable influence a broader situational context or background knowledge may have on the interpretation process. To that end, Coulson and Oakley (2005) propose the notion of a grounding box that appears to capture such more contextually-laden aspects of meaning. These authors argue for the importance of interlocutors' shared background knowledge and assumptions about the world – a one-dimensional grounding box – that help them make sense of all kinds of referential statements. Antović (2016) goes one step further by refining the grounding notion so that it encompasses different layers of meaning making, instead of representing a monolithic construct. The theory of multilevel grounding or multilevel-grounded semantics thus proposes a meaning construction model whose grounding box involves six hierarchical and partially recursive levels of interpretation: *formal*, *cross-modal*, *affective*, *conceptual*, *culturally rich*, and *individual*.

This framework arose as a way to analyse musical meaning in a systematic way, capturing the possible unconscious mechanisms that lead from the simple perception of musical structures (the first stage of multilevel grounding) toward the most complex, metaphorical interpretation of a piece enriched with the listener's personal experience (the final grounding stage). To illustrate, Antović et al. (2016), spotted a curious verbal description of a musical piece given by one of the participants. This person stated that a 20-second excerpt from Grieg's piece *In the Hall of the Mountain King* reminded them of "a fat lion from a cartoon, walking furtively, on its tiptoes" (Antović, 2021a, p.149). In an attempt to explain how such metaphorical interpretation does not simply come out of nowhere, Antović (2022) lays out the six-stage process of meaning making, highlighting the blends that appear on all levels of grounding and ultimately motivate such a highly personalized interpretation. On the first level of meaning generation, the author argues, the insightful participant unconsciously picks up the formal aspects of the musical piece they have been exposed to: in this case, the legato and staccato segments with slow crescendo motivating the overall perceptual tension. The second level is where the participant draws parallels between the formally observed structures and their own bodily experience, resulting in the staccato tones creating a PATH image schema in the mind of the listener, that are ulti-

mately going to be perceived as 'steps'. Level three, or the affective tier, attaches to the previous interpretation a comic sensation based on the idea that the steps so perceived seem heavy yet purposefully chopped up and quieted down. On the fourth level, basic entities performing the action are born, such as humans or animals sneaking so as not to get noticed. The higher, fifth level of grounding calls for the listener's cultural knowledge, now creating a specified and contextually appropriate blend: the funny entity traversing the path on its tiptoes must be a fat lion often seen in cartoons, its robust posture juxtaposed against its attempt to go unnoticed. Lastly, on the personal grounding tier, the same participant might have had a recollection of something similar happening in real life – a large animal trying to scurry its way out of trouble. The final level of grounding is the one most susceptible to personal experience, which is why the same composition could, in the mind of another listener, trigger a slightly different interpretation – for example, a fan of Steve Cutts' *Man* might, upon hearing the same piece, imagine a human rather than an animal scurrying their way out of the mess they always seem to leave behind. (The latter interpretation involves a more sombre overall sensation when the sixth-level association of the musical piece to the famous short film calls for a less funny and positive interpretation on the third, affective level, which explains the partial recursiveness of the proposed grounding process.)

So far, multilevel grounding has provided proposals as to how to uncover the relevant meaning-making operations in the domains of music, visual art, and language (Antović, 2016, 2021a, 2021b). However, it has done so in one-dimensional input channels alone, without analysing data that incites more than one sensory organ at the same time. A broader question this paper thus poses is: *What happens when overlapping linguistic and auditory stimuli appear to be sending out contradictory signals to the meaning-making apparatus inside our minds?* In other words: *How is it that we interpret incongruous data coming from the domain of language and sound – do we arrive at meaning by completely rejecting one stimulus or do the two conflicting stimuli work together to generate novel (and often unexpected) layers of meaning?*

The present paper advocates for the second option, drawing on the insights provided by the theories of multilevel grounding and conceptual blending.

More precisely, this paper provides multilevel-grounding analyses of five scenes taken from popular TV shows featuring deadpan humour (*Westworld*, *Stranger Things*, *Breaking Bad*, *Killing Eve*, and *Wednesday*). The analyses through six layers of grounding are meant to pinpoint the exact levels on which the music-language incongruities happen and ultimately provide an explanation of how these two communication modes appear to misalign with one another in order to introduce novel meanings. Following the first, *introduction* section that covers the theo-

retical framework against which the analyses are carried out, section two outlines the *methodology* used, while section three (*analysis*) presents the detailed analyses of five TV show excerpts featuring 'mismatches' in dialogue and background music. Section four or the *conclusion* part revisits these insights and outlines some ideas for further research.

Theoretical Framework

The question of how the human conceptual system interprets different kinds of data by ascribing meaning to them has been one of the most basic topics in cognitive science. More recent theories of meaning that arose from the school of cognitive semantics treat language as part of a conceptual system operating at an unconscious level, not functionally separated from other cognitive domains like memory, perception, attention, etc. Cognitive phenomena exhibited by language such as conceptual metaphor can thus also be found in visual and auditory perception (for evolutionary similarities between language and music with respect to recursion, group membership, and the expansion of connotations, see Antović, 2009). Lakoff and Johnson (1980) base their well-known metaphor proposal on the notion of cross-domain mappings that connect the source domain (a more 'familiar', concrete, and embodied conceptual packet from which one draws the metaphor) to the target domain (a more abstract domain to which one applies the metaphor). This explains why we often refer to arguments using war terminology (e.g. *His criticisms were right on target*), why heaven occupies the upmost position in *The Ladder of Divine Ascent* icon or Michelangelo's *Last Judgment* fresco, or why even instrumental sections of high-pitched songs like *Walking on Sunshine* by Katrina and the Waves, arguably, sound happy. The proper interpretation of two metaphors in question, ARGUMENT IS WAR and UP IS GOOD, DOWN IS BAD, is made possible due to the process of embodiment and Mark Johnson's (1987) image-schemas, abstract cognitive constructions arising from our embodied experience that help us make sense of the world around us. 'On target' thus relies on the image schema known as CENTRE/PERIPHERY and the metaphor rooted in our biology that presupposes that central parts are usually more important than peripheral ones. Therefore, the metaphorical sentence from above roughly translates to: *His criticisms were accurate and relevant*. The other two examples coming from the domain of visual and auditory perception instantiate the image schema known as VERTICALITY which, once again, makes use of our embodied experiences to convey the meaning that things positioned high on a vertical scale are essentially better than the lower ones.

An important extension to the conceptual metaphor theory in cognitive science, focusing more on 'emergent', 'online' aspects of meaning generation, is known as the theory of conceptual blending (Fauconnier & Turner, 2002). To explain the essence of this approach, one may start

with an example sentence: *If Bill Clinton had been French, there would not have been the Monica Lewinsky affair* (Antović, 2021a) – a paraphrase of Coulson and Oakley's (2005): *In France, the Lewinsky affair wouldn't have hurt Clinton.* The idea of Clinton being French blends two mental spaces (that of an American president and his counterpart in France), mapping certain characteristics of each one onto the other (say, their political careers and susceptibility to affairs) that ultimately results in a hypothetical scenario in which such an affair would not have been dangerous to presidency. The ability to understand this counterfactual sentence stems from the interlocutors' shared knowledge about the American president, his personal life, and the political system in France – all the things that would go into Coulson and Oakley's (2005) grounding box. Similarly, a recent meme that has been circulating on social media since the premiere of Christopher Nolan's *Oppenheimer* shows a scene featuring the characters of Robert Oppenheimer and Albert Einstein, while the caption reads: *this must be like The Avengers for people who know multiplication tables*. To get the joke, one cannot simply map the domain of 'people who know multiplication tables' onto 'regular population that attends blockbuster premieres' and the domain of '*The Avengers*' onto the domain of '*Oppenheimer*'. Instead, there are certain layers of meaning at play here that cannot be accounted for only in terms of cross-domain mappings that conceptual metaphor theory proposes. The fact that the joke is meant to 'offend' people who know multiplication tables by implying that they are individuals obsessed with mathematics who get excited about scientists instead of superheroes and also imply that the people who know multiplication tables make up a minority means that there is much more going on here than the conceptual metaphor framework would be able to explain. The theory of conceptual blending, on the other hand, would call upon the notion of mental spaces for the analysis – small cognitive packets constructed during discourse that help interlocutors organise information in order to make sense of it. This theory would thus recognise two input spaces in the meme: the mental space of people who know multiplication tables (smart people who look up to scientists and probably do not visit cinemas as much as Marvel fans do) and the mental space of regular individuals (who are not as familiar with basic mathematical operations and rarely miss an event in the Marvel Cinematic Universe). The humorous effect of the meme is achieved by selectively projecting the relevant elements of the two input spaces together into a blended space so as to generate novel meaning.

A useful addition to the theory of mental spaces and conceptual blending comes from Coulson and Oakley (2005). In examining the relationship between literal and figurative meanings, these authors borrow Langacker's (2002) notion of *ground* that includes the speech event, participants, and context and arrive at the conclusion that it should be ex-

panded as to specify participants and roles, forum, and context surrounding the utterance. Antović (2021a) introduces 'multilevel' grounding, claiming that the grounding box should not constitute a single, monolithic square in the conceptual blending diagram. Rather, it should be understood as a multilayered construct, composed of six partially recursive and hierarchical grounding levels: formal, cross-modal, affective, conceptual, culturally rich, and personal. On the first, *formal* or *perceptual* level, the human conceptual system parses stimuli into logical (linguistic, musical, or visual) chunks that further get schematized on the second, *cross-modal* level by undergoing processes motivated by Johnson's image schemas. On the third, *affective* level, the previous interpretation gets assigned emotional valence that results from the combined formal and cross-modal experience. On the fourth, *conceptual* tier, very basic narratives are constructed, while on levels five and six, *culturally rich* and *individual*, these concepts are coloured with cultural knowledge and personal experience. What is important for the analysis is that higher grounding tiers always build upon the interpretation coming from the lower ones.

To illustrate further, let us outline an example: a participant's interpretation of Wagner's sword motif in *Rheingold* as "the descent of gods from Olympus" (Antović, 2021a, p.152). A multilevel-grounding analysis of the piece suggests the participant's perception of tension arguably stemming from the formal characteristics of the composition on level one: a clean C major triad laid against a broken chord that ultimately disappoints the listeners' expectancies. Further blended with the loudness and abruptness of ascending pitches, a sense of movement emerges, which, in combination with the prevalent forte dynamics, gives rise to the FORCE schema on the second, cross-modal layer. The so perceived loud 'clash of forces' with alternating tension and relaxation results in 'drama' on the affective level. This level-three blend then integrates with additional elements inferred from the musical construct, such as the loud baritone voice and the blazing brass section, to evoke battle imagery on level four, further specifying the army so imagined as victorious due to the 'conventionally majestic' C major. On level five, the participant checks the constructed military narrative against their cultural knowledge of Greek mythology, finally arriving at the 'ancient Greek divinities descending from Olympus'. Such an understanding can further be coloured with one's own associations to god(s), battles, or ancient Greece, allowing for a variety of possible interpretations on the final grounding level.

Returning to the *Oppenheimer* example from above, our meaning-making apparatus would first 'grammatically' parse the linguistic and visual stimuli in the meme into clusters around heads of phrases (in language) and constitutive elements of figures/objects (in vision), schematize them, and attach emotion to the words and components of the image. Only on level four would the caption start to assume a more tractable, per-

haps conceptual meaning, but it would still be very basic. In other words, our meaning-making apparatus, upon attaching higher importance to certain words and phrases on level three (possibly the ones that would turn out to be more referential, fuller in the sense of having lexical rather than grammatical meaning, and more stressed in pronunciation, such as 'Avengers' and 'multiplication tables'), would now begin to motivate a simple narrative on level four. The fourth-tier analysis would first link the word 'this' to the image it probably refers to and linger over the two referential words in question and the simile-introducing word 'like', evoking the very basic meanings of the things or entities being compared. And, while everyone might arrive at a conclusion that the image depicts two people standing (and possibly conversing) by the lake (still level four), some people might find it difficult to get the joke as their (proper) interpretation of the meme would likely end there. This highlights the importance of shared cultural knowledge made evident no earlier than on level five. Recognising the character of Einstein might also come naturally to most people, but what about the Oppenheimer character? As Cillian Murphy's face is not clearly visible in the picture, some might struggle with identifying both the actor and the character he portrays. Again, for those who have not even heard of the film, it would be impossible to allocate the image to the '*Oppenheimer* movie' mental space upon seeing the way the two characters are dressed: this would require shared background knowledge existing only on level five or the cultural tier. The same goes for language – to some, the word 'Avengers' might not evoke the superheroes from the Marvel Cinematic Universe at all. Thus, it may not build an 'MCU event premiere' mental space, as proposed above, such that most folks following this genre never miss. Returning to the image, those who have seen the movie or are familiar with the lives of the two scientists in question might consider the joke even more on point after realizing the analogy between the tension that exists between Marvel protagonists and antagonists on the one hand and Einstein and Oppenheimer on the other. Finally, on the personal level, that conflict might further resemble a recent argument with a co-worker or a disagreement with a friend or partner that we ourselves had. On the individual level, some people might also get offended for knowing multiplication tables and the fact that the meme is meant to portray them as 'geeks', while others may feel embarrassed for potentially *not* knowing anything about more advanced calculus.

Multilevel-grounded semantics has so far proven useful in several studies. It has been considered as a methodological alternative alongside conceptual integration in humour studies of short stories published in Italo Calvino's *Cosmicomics* collection (Marković, 2022a). The two allegorical stories the paper analyses (*All at One Point* and *How Much Shall We Bet*), set in a pre-Big Bang era, call for the immersion of the reader into the storyworld lacking the time and space as we now know them. The

author extends the idea that this immersion can be seen as the process of constructing macroblends, suggesting that such blends are formed through a series of successive mappings delineated by different levels of constraint. A comic interpretation thus comes as a result of blends occurring on different grounding levels: e.g., the allegory of all beings stuck in an enclosed space before the Universe was created with too much time on their hands when time did not even exist disappoints expectancies and leads to a humorous effect on the formal and cross-modal grounding level by implying a simultaneous existence of space and time and their negation. Unlike the original Italian and English version, the Serbian translation of Calvino's short story also mocks the readers' attempt to completely immerse into the storyworld on the culturally rich grounding level when the narrator, lamenting over the narrowness of their living situation, uses the vocabulary that implicitly references hell and devil (or paradise and God) in a world that still lacks both. Apart from applying the multilevel-grounding framework to humour analysis in longer narrative texts, the same author (Marković, 2022b), analyses how the meaning of 'the grotesque' varied throughout the poetic discourses of Pre-Romanticism and Romanticism. She suggests the multilevel-grounding model be applied in future studies dealing with the elements brought into grotesque connection. Finally, the pending study by Havlíčková Kysová (2024) adopts the same methodological framework as the present paper as a possible analytical approach to operatic production. Other authors, including Bonifazi (2024) and Patrick et al. (2023), have also implied the potential applicability of the multilevel-grounding model in their work.

METHODS

Aim and Hypothesis

Given all the layers of meaning the theory of multilevel grounding can expose, it has proposed some potential novelties in the semantics of music, visual art, the language of poetry, stock phrases, grammatical constructions, and idioms (Antović, 2016, 2021a, 2021b). The present paper is meant to take this one step further and consider in greater detail what happens during the process of interpreting *conflicting* musical and linguistic information. This study thus poses the following research question: *During the meaning-making process, when our conceptual system is presented with incongruous data coming from the domain of music and language, do we completely reject the interpretation of one of the stimuli or do the two stimuli still combine to form novel layers of meaning?* The hypothesis of this paper is that the *inputs that appear to be contradictory on some grounding levels still work together to generate novel and often humorous or unexpected meanings.*

Sample

In order to pinpoint the exact spots where meaning incongruities happen, this study carries out a multilevel-grounding analysis on a sample of five scenes taken from popular TV shows in which the characters' dialogue and the background music seem to be sending out contradictory signals on one or some of the grounding levels this theory presupposes – *Westworld* (season 02, episode 07, at 13:51 until 14:38), *Stranger Things* (season 01, episode 03, at 39:03 until 41:43 and at 45:11 until 45:58) *Breaking Bad* (season 04, episode 13, at 41:20 until 44:13), *Killing Eve* (season 03, episode 02, at 27:19 until 29:29), and *Wednesday* (season 01, episode 01, at 27:24 until 29 min).

ANALYSIS

For each scene, qualitative analysis is carried out so that it highlights the mismatches that exist between music and dialogue, marking the exact grounding levels on which these incongruities occur.

The **first scene** is taken from Netflix's *Stranger Things*, a popular TV show that combines the '80s nostalgia with elements of science-fiction. The Christmas lights scene in question features the character of Joyce (Winona Ryder) and displays her attempt to communicate with her missing son, Will, stuck in the parallel world called *The Upside Down*. Due to the atypical nature of the interaction going on in this scene, it does not represent a proper dialogue – while Joyce carries the message across using words, Will, who is not *truly present*, communicates by turning the lamps and Christmas lights around the house on and off. Music-wise, this scene features an instrumental piece called *One Blink for Yes* by Kyle Dixon and Michael Stain.

Our analysis of the scene starts on level one of the presupposed grounding box: perceptually, the cognitive system first recognises instances of language, on the one hand, and background music, on the other. In terms of language, there is no true, fluent speech – the words coming from Joyce's mouth can only be characterised as hesitant speech with fillers like 'um', breaks, and stutter. Because of that, formally, Joyce's speech does not have a final point and lacks a climax (though, if we take into consideration the visual component, there *is* a linguistic climax once the word 'RUN' - her son's warning from 'the other side' - is spelled out one letter at a time using Christmas lights, at the very end of the scene). On this level, the cognitive system still parses the linguistic elements that will be necessary for further interpretation – separating, for instance, 'the lightbulb spelling used by Will' from the rest of the context, or Joyce's character from the visual background. On the second, cross-modal tier, we immediately infer the BLOCKAGE image schema, and possibly CON-

TAINMENT that stem from constant breaks and hesitation noticeable in Joyce's speech that seem to prevent her from proceeding, both speech and movement-wise. On the other hand, the tones heard in the background formally appear pulsating (level one), and the music reaches a peak towards the end of the scene (level two). Cross-modally, this creates the PATH image schema, with a discernible beginning and end. On the affective level, language-wise, we might observe phrases that, purely aurally, suggest optimism, at times emerging from the generally sombre words, which once again accounts for the *partial* recursiveness of grounding levels in that the BLOCKAGE schema from level two does not necessarily affect the optimistic tone emerging on level three. When it comes to music, the pulsating tones played in a minor scale, together with the path schema that creates a feeling of anticipation, give rise to a sense of sinister foreboding so that the entire composition sounds rather eerie. On the fourth, conceptual level, the interrupted language and the BLOCKAGE schema together with a need for hope that stems from Joyce's (at times) optimistic tone, create the feeling of an object/person stuck somewhere. On the culturally rich level, this kind of scenario gets refined in that *Stranger Things* fans might imagine Will being lost in a fictional world or Joyce equally trapped in the real world. Even higher along the grounding plane, an individual might further think of a concept of being stuck and imagine a scenario they faced an unpleasant situation they could not solve or escape from, some kind of a personal predicament. Returning to music, the sinister minor-scale tones that resemble an object traversing a path pave the way for a simple scenario on the conceptual level: we might imagine a person cautiously walking or travelling from point A to point B. On a higher, cultural level, this might be understood as Joyce apprehensively walking down the imaginary road that leads to her son. Finally, from a personal angle, this analysis might further evoke a particular instance of travelling or moving along a path with a sense of unease. What can be deduced from the parallel analyses of linguistic and musical data in the scene is that they differ on all five levels of grounding (the individual level being excluded from the comparison on purpose due to the lack of objectivity such interpretations would entail). The perceived mismatches between music and language throughout the scene might serve as a foreshadowing of the unexpected end, highlighting the fact that, at this point, Joyce is nowhere near finding her son. This way, the PATH image schema we recognise in music 'clashes' with the BLOCKAGE schema in language, but does not completely override it. Instead, the two schemas work together to communicate the idea that Joyce still has a long way to go (and many obstacles to pass) to reunite with Will.

The **second scene** this paper analyses is taken from an AMC show called *Breaking Bad*. The show is an absurd depiction of a gradual transition of a high school chemistry teacher, Walt (Bryan Cranston), diag-

nosed with terminal lung cancer, into a meth producer and dealer. Many scenes throughout the show rely on dark humour that often arises from unexpected situations and absurdity. The scene in question happens when Walt and his partner, Jesse (Aaron Paul), manage to get rid of their boss and are finally free to gain more independence and conduct their business the way they want. The dialogue in the scene is scarce, with only Walt taking his turn in speaking, i.e. conveying information to Jesse in an attempt to organise their business. In terms of music, this scene includes an instrumental piece called *Freestyle* by Taalbi Brothers.

Looking at language alone, we can see that it is used minimally, as it consists of only three short lines (*Gus is dead./We've got work to do./Vamonos.*), though we might still say it formally reaches a peak at Walt's final utterance because this line is perceptually different from the other two: it comes after a longer break, its tone/affect is different, and finally, unlike the first two utterances, it is in Spanish. The three short lines Walt utters are separated with breaks. The second of the two is longer and additionally introduces the linguistic peak – Walt's final utterance (*Vamonos*). More importantly, the tension seems highest at the beginning of the scene, with Walt's disclosure of crucial information regarding the death of their boss, and slowly relaxes towards the end, again, with the 'Vamonos' line, which suggests that the partners have suddenly obtained more freedom in business. With regard to music used in this scene, we notice its incongruence with language even on the first, perceptual level of grounding: unlike the language, the music is repetitive and continuous – we can hear the upbeat guitar tune that reaches a peak towards the end of the scene, matching the peak we witness in language. Cross-modally, the lines prior to the linguistic peak are monotonous and static and do not instantiate a clear image schema. On the third, affective level, the two blends combine with the serious and determined tone of what is being said. The music, however, is fast-paced with interrupted LINKS on the cross-modal level and played in a major key, which, on the affective level, further gives rise to a rather careless and optimistic scenario. Also enhancing this interpretation one finds the fast tempo of the tune, which motivates excitement. Conceptually, Walt's lines create an image of a person performing a difficult task, his three short utterances evoking concepts such as unpleasantness and difficulty, but also determination and resourcefulness. When it comes to music, the previous-level blend now evokes an exotic and lively scene rather than a critical situation, and we might imagine a person running or moving vigorously on the conceptual grounding level. Only on the culturally-rich level, however, do we understand that Walt's words in fact refer to the death of their boss – upon ensuring their boss is dead, Walt and Jesse are now ready to burn down their meth lab. Musically, on the same level, the vigorous movement might further become a Spanish flamenco dance evoked by a fast-paced tune re-

sembling a rhythmic stamping (zapateado) that, once again, does not have much in common with the severity of Walt and Jesse's previous (and future) actions. Instead, such deliberate discrepancies that emerge on all five levels of grounding only add to the absurdity of the main characters' situation. Similarly to the *Stranger Things* scene from above, the (lively) music that does not give out the same feeling as the words uttered serves to highlight all the ridiculous actions Walt and Jesse have taken to get to this point and to anticipate the same chain of events re-occurring in the future.

The **third scene** we analyse comes from a BBC America show called *Killing Eve*, a dark comedy that revolves around childish Villanelle (Jodie Comer), a hired assassin, constantly chased by MI5 agent Eve Polastri. The show features a lot of witty scenes that make use of gallows humour and an unexpected turn of events. In the scene in question, the impatient Villanelle embarks on an assassination mission with Felix, an utterly inexperienced newbie. The language used in the scene consists of a series of short lines rich in irony. Musically, the scene includes a piece called *Danger* by Unloved.

A possible linguistic analysis starts on level one where we formally perceive numerous short lines that seem to reach a peak at the end of the scene, with Villanelle's final punchline: *Management sucks*. The prosodic units seem evenly distributed until the longest pause leading to Villanelle's remark – *What happened to the plan?* – where the tension suddenly appears to rise, decreasing only after the final punchline at the very end of the scene. The level one music analysis, on the other hand, reveals dissonance with a quite ominous undertone and the final few notes lacking, which directly contradicts the linguistic peak at the end of the scene. On the second, cross-modal grounding level, we notice a slow, linear movement toward the said punchline in language. In music, we can identify the CYCLE image schema that originates from a whirlpool-like movement of tones, which is once again incongruous with the linearity of Villanelle's intonation. On level three, a touch of seriousness is given to the previous linguistic analysis, stemming from Villanelle's deep voice and the fact that her intonation appears even throughout most of the scene. This further gives rise to the idea that she is making an arrangement to eliminate a target on level four. The fourth interpretation level thus evokes an image of assassination, together with the roles the two interlocutors play in it: either a hitman or the victim, or – in the case of Felix – both. The third, affective tier in music placates the interpretation by adding to it a more lively vibe (despite the lyrics), in the sense that the rhythm here grows strong and energetic prior to the line where tension suddenly rises, resulting in a sense of mockery. On the conceptual tier, this might resemble an effortful, staggering movement in circles that further evokes some kind of a psychedelic pastiche on the fifth, culturally rich level of grounding. The same, cultural tier concludes the linguistic narrative in the sense that two

professional assassins, Villanelle and Felix, fail to perform an execution neatly, drawing the viewers' attention to what is so typical of this show and most of the characters in it: nothing ever gets accomplished the way it is supposed to, which is part of the 'cultural' knowledge of *Killing Eve* fans. In a way, keen viewers become conditioned to always expect the unexpected. Again, in the eye of a careful observer, the differences that are visible on all five levels of grounding introduce the twist at the end of the scene – the fact that Villanelle, irritated with Felix's lack of subtlety in eliminating their target, finally takes matters into her own hands and kills both of them – the absurdity of the scene being on a par with many similar situations occurring in the show. Put differently, the mismatches between language and music present in this scene appear to be deliberate, serving the purpose of foreshadowing, announcing or introducing ahead of time the slaughter at the end of the scene together with the punchline – Villanelle's witty final remark.

The **fourth scene** we analyse comes from Netflix's *Wednesday*, a teenage adaptation of *The Addams Family* that follows the adjustment of Wednesday Addams (Jenna Ortega) to the new school. The show combines the original movie's eccentricity and quirkiness with the technological advancements of the 21st century, all the while being full of a morbid sense of humour and dark twists. Much like in the *Stranger Things* scene, the communication going on here is unconventional as it occurs between Wednesday using words and Thing who 'speaks' using gestures. As for the soundtrack used in the scene, it is a traditional Mexican song called *La Llorona*, performed by a female artist called Chavela Vargas.

In terms of the perceptual analysis of language structure in the scene, on the first level, we perceive longer chunks of text that contain lexically rich sentences. Musically, the first level of our analysis includes uninterrupted tones with volume strongly decreased at one point (staying low throughout most of Wednesday's speech). Cross-modally, Wednesday's lines are unvarying/linear and do not instantiate a clear image schema. On the other hand, in the musical piece played in the background, we are able to notice the VERTICALITY schema, since the tones accompanied by the performer's voice constantly fluctuate. Affectively, the lengthy monotonous monologue combines with a serious tone of what is being uttered, stemming from Wednesday's flatness of intonation. On the further, conceptual level, this generates a quite schematic context of a person being scorned or threatened. The interpretation coming from the lower two levels in music, however, combines with the romantic and mournful affect on the third level of grounding, which further gives rise to the basic concepts of courting or mourning on the conceptual tier. Finally, on the cultural level, the narrative expands to feature Wednesday's attempt to negotiate with Thing, sent by her parents to spy on her. When the analysis moves on to the culturally rich level in music, we can imag-

ine either a Spanish romance (upon hearing a fado) or an unlucky scenario of someone mourning as a result of unrequited love. As before, the discrepancies between music and language that sustain up until the fifth level of grounding seem deliberate and their aim is to introduce the final plot twist of the scene: Thing ultimately 'pledging [its] undying loyalty' to Wednesday instead of continuing to spy on her for her parents.

The **fifth scene** this paper examines is taken from an HBO show called *Westworld*, a science fiction franchise dealing with moral and ethical issues surrounding the running of an android-populated amusement park. The show abounds in unsettling scenes of newcomers (regular people) doing all kinds of horrible things to hosts (androids), but also features a lot of philosophical scenes, mainly the ones with Ford (Anthony Hopkins) – the park's founder – dealing with God and man playing God, artificial intelligence, consciousness, free will, etc. This scene tackles the issue of the true nature of the park, revealing the horrors that happen in it. The dialogue in the scene alternates between Ford and Bernard (one of the founders who turns out to be a host – one of the androids inhabiting the park). In terms of the background music played at the end of the scene, it is an adaptation of the *Westworld* theme song.

When it comes to language, on the first, perceptual level, we can notice interrupted sentences characterised by hesitation. The unchanging tone leads up to the final section that seems different from the rest of the dialogue. Thus, on the second, cross-modal level, we perceive monotonous lines that, on the affective level, get enriched with a sombre sentiment, while the more dynamic utterances from the final part of the speech (level two) result in a lighter, more positive affect on level three towards the end of the scene. In terms of music, on the formal level, we can identify uninterrupted legato tones during most of the scene that ultimately lead to staccato tones towards the end. This gives rise to the LINK schema (during the legato section), clashing with interrupted speech and turn-taking in language, and the lack of it (staccato), as well as the PATH and VERTICALITY image schemas on the cross-modal tier, which again misaligns with the linearity of the dialogue on the second level (Antović, 2018). On level three, the interpretation is coloured with a more lively sentiment that arises from tones played in a major chord at the end of the scene. On the fourth level of grounding, the linguistic analysis might reveal a basic narrative such as speaking of death, destruction, and creation. At the same time, the previous-level blend in music further gives rise to the image of a river flow (for the PATH schema) or hopping (for the VERTICALITY schema). Finally, on the fifth, cultural level, the conversation anticipates the talk about the true nature of the park and evokes the religious implications that are present throughout the show. With regard to music, the previous images might evoke an even more elaborate scene of dancing on the final level of grounding. The perceived differences between language and background music,

once again, serve to introduce the subsequent revelation of the amusement park's true nature, highlighting and even mocking the illusion its residents seem to have been under all this time.

CONCLUSION

The present paper hopes to have taken a small step in extending the multilevel grounding approach to the analysis of multimodal data that combine different sensory information at the same time. More precisely, we have advocated in favour of a multilevel grounded semantics approach to revealing different layers of meaning that may sometimes remain uncovered by less contextually-driven approaches. The theory of multilevel grounding, thus, may help in the process of dissecting not only one-dimensional inputs such as linguistic, visual, or auditory information (Antović, 2016, 2021a, 2021b), but also combinations of stimuli coming from different modes. In particular, by taking a complex sample such as a TV show scene, this paper has aimed to reveal what happens when two stimuli appear to be sending conflicting signals to the meaning-making system of the viewer. Multilevel grounding analyses of five scenes taken from popular shows featuring deadpan humour and absurdity (1) reveal how this approach seems suitable for explaining the process of meaning construction on each mode separately, (2) pinpoint the exact contextual layers where incongruities between linguistic and auditory stimuli happen, and finally (3) account for the novel and often unexpected meanings that arise this way, such as plot twists that lead to often humorous yet always non-composite interpretations. However small this novel meaning might be, the observed mismatches between language and music typically serve well to explain its motivation. Put differently, while some instances of meaning might be deduced from language or music alone, it is always the combination of the two – more importantly, the deliberate conflict between the two – that adds this new layer to the standard interpretation (e.g., the sudden anticipation of an unpleasant event in the *Stranger Things* scene, falling out of 'business partners' and the unexpected death of one of them in *Killing Eve*, foreshadowing of equally bizarre future actions of Walt and Jesse in *Breaking Bad,* the slow announcement of Thing suddenly, though unwillingly, taking Wednesday's side in *Wednesday,* and the shocking revelation of the true nature of the amusement park that follows in *Westworld*). Finally, the multilevel grounding framework seems to better account for such musico-linguistic mismatches than other theories dealing with the phenomenon of (disappointed) expectancy in that it disassembles each (incongruent) stimulus into its constituent parts – starting from its more easily observable formal characteristics to the grounded personal associations they motivate – and compares it with the same properties of the other stimulus dissected in the same way. This ul-

timately makes the comparison of incongruous stimuli more systematic and detailed, possibly leading to a more reliable overall analysis.

Future research in this field might include similar studies that would feature a larger sample or different kinds of samples containing contrasting language and music, and might take into account data coming from the domain of vision as well. Potential studies might also rely on different participants' analyses of such stimuli that would engage their senses in distinctive ways so as to see how this might affect the comprehension process.

REFERENCES

Antović, M. (2022). *Multilevel grounding: A theory of musical meaning*. Routledge.

Antović, M. (2021a). Multilevel grounded semantics across cognitive modalities: Music, vision, poetry. *Language and Literature*, *30*(2), 147-173. https://doi.org/10.1177/0963947021999182

Antović, M. (2021b). Contrasting English and Serbian through the lenses of multilevel grounded semantics. *Nasleđe, 48*, 165-181.

Antović, M. (2018). Schemas, grounds, meaning: On the emergence of musical concepts through conceptual blending. *Musicae scientiae*, *22*(1), 57-71. https://doi.org/10.1177/1029864917711218

Antović, M. (2016). From expectation to concepts: Toward multilevel grounding in musical semantics. *Cognitive Semiotics*, *9*(2), 105-138. https://doi.org/10.1515/cogsem-2016-0005

Antović, M. (2009). Za i protiv neodarvinizma: primer jezika i muzike. (For and against Neo-Darwinism: the case of language and music.) *Teme*, *33*(4), 1169-1190.

Antović, M., & Tasić, M. (2011). Muzika na granici i teorija konceptualne integracije: analiza jedne pesme sa juga Srbije. (Music on the border and conceptual integration theory: analysis of a song from the south of Serbia.) *Teme, 35*(4), 1521-1542.

Antović, M., Stamenković, D., & Figar, V. (2016). Association of meaning in program music: On denotation, inherence, and onomatopoeia. *Music Perception, 34*(2), 243–248. https://doi.org/10.1525/mp.2016.34.2.243

Bonifazi, A. (2024). Cross-modal iconicity in songs about weeping. *Semiotica*, *2024*(256), 1-29.

Coulson, S., & Oakley, T. (2005). Blending and coded meaning: Literal and figurative meaning in cognitive semantics. *Journal of pragmatics*, *37*(10), 1510-1536. https://doi.org/10.1016/j.pragma.2004.09.010

Faucconier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.

Havlíčková Kysová, Š. (2024, October 10-11). *Multilevel grounding theory as analytical approach to operatic production – theatre studies perspective* [Conference presentation abstract]. Meaning of Music, Academy of Music in Gdańsk, Poland.

Johnson, M. (2005). The philosophical significance of image schemas. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: Image schemas in cognitive linguistics*, Vol.29, 15-33. Berlin, Germany: Walter de Gruyter.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason.* University of Chicago Press.

Lakoff, G. (2006). Conceptual metaphor. In D. Geeraerts (Eds.) *Cognitive linguistics: Basic readings*, *34*, 185. Walter de Gruyter.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* University of Chicago, Chicago, IL.

Langacker, R. W. (2002). *Concept, image, and symbol: The cognitive bases of grammar.* Walter de Gruyter Inc.

Marković, O. (2022a). Kako razumemo priče smeštene u period pre velikog praska? Nivoi utemeljenja kao metodološka alternativa pri proučavanju humora. (How do we understand stories set in the period before the Big Bang? Levels of grounding as the methodological alternative in humor study.) U: Mišić Ilić, B. & Lopičić, V. (Ured.), *Jezik, književnost, alternative* (str. 329-343).

Marković, O. (2022b). Groteska kao kognitivni fenomen poetičkog diskursa u predromantizmu i romantizmu. (Grotesque as a cognitive phenomenon of poetic discourses during Pre-Romanticism and Romanticism.) *Philologia Mediana*, *14*(14), 97-111.

Patrick, M. T., Cohn, N., Mertus, J., & Blumstein, S. E. (2023). Sequences in harmony: Cognitive interactions between musical and visual narrative structure. Acta Psychologica, 238, 103981.

МУЗИКА И ДИЈАЛОГ СУКОБЉЕНИ У СЦЕНАМА ИЗ ПОПУЛАРНИХ ТВ СЕРИЈА КРОЗ ПРИЗМУ СЕМАНТИКЕ ВИШЕСЛОЈНОГ УТЕМЕЉЕЊА

Ања Тошић, Михаило Антовић

Универзитет у Нишу, Филозофски факултет, Ниш, Србија

Резиме

Након вишедеценијског развоја формализма у семантици који је настао на основу учења Ноама Чомског, нешто новије семантичке теорије донеле су доста промена у погледу језичких феномена који ће се током даљих година изучавати. Највећи скок у том смеру направила је теорија појмовне метафоре Џорџа Лејкофа и Марка Џонсона, препознајући исти феномен како у језику, тако и у музици и визуелној уметности, што даље указује на постојање метафоре на вишем, појмовном нивоу, односно нивоу мисли. Сличну идеју видимо и код Фоконијеа и Тарнера у теорији појмовног сажимања којом се објашњава стварање новог значења на основу пресликавања одређених елемената између два или више ментална простора. С тим у вези, аутори Шона Кулсон и Тод Оукли предлажу појам кутије утемељења којим би се обухватило комплетно позадинско знање које саговорници треба да поседују како би таква врста пресликавања, и на крају споразумевање, били могући.

Теоријом вишеслојног утемељења, Антовић усавршава овај појам и, полазећи од тога да она не треба бити једнодимензионална, даље дели кутију утемељења на шест хијерархијских и делимично рекурзивних нивоа: формални ниво на коме се језички, музички или визуелни елементи деле на целине на основу заједничких карактеристика, кросмодални ниво где долази до јављања преконцептуалне схеме попут Џонсонових сликовних схема путање, вертикалности, центра/периферије, афективни ниво који у претходну интерпретацију уноси емоционалну вредност, појмовни ниво који је одговоран за формирање простих наратива, културолошки ниво на коме се наратив обогаћује позадинским и контекстуалним знањем, и, на крају, индивидуални ниво на коме се, на основу свих претходних нивоа, формира лични доживљај.

Како је теорија вишеслојног утемељења нашла примену у анализи развоја значења како у језику (поезији, идиомима и фиксним фразама), тако и у музици и визуелној уметности појединачно, у овом раду она служи као методолошки оквир за испитивање начина на који се долази до разумевања музичко-језичког наратива у пет сцена из ТВ серија у којима се дијалог и позадинска музика сукобљавају на једном или више нивоа утемељења које ова теорија претпоставља. Појединачном анализом језичких и музичких података из серија Западни свет, Чудније ствари, Чиста хемија, Убити Ив и Среда, овај рад открива на којим тачно нивоима долази до неподударања између језика и музике, сугеришући да управо овакве разлике и доводе до духовитог или изненађујућег ефекта на крају сваке од датих сцена.