Dynamic Schematic Complexes:

Image Schema Interaction in Music and Language Cognition Reveals a Potential for Computational Affect Detection

Mihailo Antović[1], Vladimir Ž. Jovanović[1], Vladimir Figar[1]

Faculty of Philosophy, University of Niš, Serbia

Correspondence concerning this article should be addressed to: Mihailo Antović, Faculty of Philosophy, Ćirila i Metodija 2, office 418, 18000 Niš, Serbia; Email: mihailo.antovic@filfak.ni.ac.rs

Abstract

The article focuses on meaning generation in music and language cognition based on complexes of image schemas. We hypothesize that schemas are involved in dynamic interactions fostering online complex conceptual construction found both in language and elsewhere in various cognitive code expressions. Expanding on recent work, we propose that an account of image-schema interaction should include the notion of *scalarity*, a higher-order parameter which captures the intensity with which schemas occur and/or interact with one another. In that sense, schemas are proposed to have a binary valence (positive or negative) and a three-tier range of intensity levels. We focus on some examples of the interaction of FORCE, PATH, LINK, BALANCE, and CONTAINMENT schemas in language and music to illustrate how a formal routine emerging from our proposal may help in addressing not only aspects of propositional semantics, but also affective and connotative nuances in musical and linguistic expression.

Key words: image schema, dynamic interaction, concept generation, scalarity, music, language.

Dynamic Schematic Complexes:

Image Schema Interaction in Music and Language Cognition Reveals a Potential for the Detection of Affective Meaning

1. Introduction

In this concept paper we present a system for capturing affective nuances of meaning in language and music based on the identification and scalarization of interacting image schemas (Hedblom et al 2015; Johnson 1987; Oakley 2007). Image schemas are relatively self-sufficient and delineable preconceptual building blocks (Popova 2005), abstract elements for concept formation in the human mind, such as PATH, LINK, or CONTAINER. When employed in metaphors, they tend to motivate expressions exhibiting emotional load, such as *Our relationship is heading nowhere* (instead of simply ending), *These two problems go hand in hand* (instead of just being linked) or *He got completely subsumed by hatred* (rather than just hating someone) (Lakoff & Johnson 1980). In the first metaphorical expression, the source domain that is mapped onto the target domain LOVE/RELATIONSHIP is DIRECTED MOVEMENT IN SPACE, which represents a metaphorical extension of the PATH schema. In the second instance, the target domain of THEMATIC CONNECTEDNESS is based on the PHYSICAL CONTACT BETWEEN TWO HANDS, and this in turn derives from the more rudimentary, concept-building schema of LINK. In the third, the overarching HATRED is based on a PHYSICAL LIQUID covering the agent from all sides: this in turn calls for the image schema of CONTAINMENT. Aiming at a possible formal detection of such connotations, we have two interrelated objectives. (1) Using music and language as test cases, we argue in favor of the idea that image schemas can engage in mutual dynamic, yet formally tractable interactions, revealing a similar abstract structure beneath apparently very different expressions across cognitive modes. For instance, the mental simulation evoked by the linguistic *car hitting a dog* and the musical sensation that the player is *hitting the peak of the tune* can be shown to have an underlying structure based on the interaction of at least two image schemas: FORCE, which produces the linguistic or musical motion, and PATH, which provides a trajectory for this motion to occur in time. (2) In addition, our sketch of a possible formal account of such image-schema interaction ("grouping", Cienki 1997; "image-schema families", Hedblom et al 2016), enriches schemas with the higher-order factor of *scalarity*, where intensities of individual schemas in the group may vary in real time. Thus, for example, the FORCE in expressions above needs to slowly increase before hitting a musical or conceptual peak. Such an enhanced analysis may help formally approach not only the denotational semantics of simple linguistic expressions, but also affective and connotational aspects of meaning in music and language: based on our analyses, we hypothesize that the stronger the individual schemas in the schematic structure, the greater the chance that the resulting interpretation will be emotionally laden, and ultimately metaphorical. After providing a bit of theoretical background (2), we use portions of the score for Beethoven's Piano Sonata No. 1 and select examples from a preliminary collection of 100 recent newspaper headlines, to argue for a dynamic approach to meaning generation in music and language (3) and to show how a formal account of scalarized image schemas in interaction could best inform such an approach (4). We then analyze some specific linguistic examples from our current corpus (5). In the conclusions (6) we stress the potential strengths and weaknesses of our analysis, adding a few words on planned further work.

2. Theoretical Background and Previous Research

Image schemas, as theoretically conceived of in the contributions of both linguists and psychologists, namely Cienki (1997), Hedblom et al (2015; 2016, 2019), Johnson (1987; 2005; 2007), Lakoff (1987, 1990), Lakoff and Turner (1989), Mandler (1992; 2004; 2012; 2014), Oakley (2007) and others, refer to specific subconscious patterns in the mind primarily based on our experience with spatial relations such as *in–out*, *up–down*, *near–far*, but also of object manipulation and movement through space. They are defined as “a recurrent pattern, shape, and regularity in, or of, [the] ongoing ordering activities…” (Johnson 1987: 29).

The source of image schemas is our immediate sensorimotor and perceptual experience coming from the interaction with the outside world, but they are not purely perceptual as they serve to build consciously accessible concepts (Mandler 2004). They start developing in the brain as early as our childhood with the distinction between *me* and *not-me* and continue throughout our mental formation processes, providing us with the meaningful interpretation essentials both for the experiential domain of the real world and the reality that exists in our cognition and communication. Different classifications of schemas have been offered in the literature, e.g., a most comprehensive one provided by Clausner & Croft (1999:15) and enriched by Evans & Green (2006: 190), with a contingent of 41 schematic patterns grouped into eight categories: Balance, Containment, Existence, Force, Identity, Locomotion, Space and Unity/Multiplicity. However, as can be verified by this and other similar classifications, the image schema SCALE often stands out from the others, resisting attempts at proper grouping. A possible reason, as we shall claim in the argument that follows, is that it is not really a schema, but rather a higher-order principle specifying the intensity with which other schemas interact.

Image schemas are always defined as dynamic, yet the range of this dynamicity is often difficult to determine. On the one end, schemas such as VERTICALITY, from the Space group, tend to denote more stable relations, while those such as CYCLE, from the Existence group, focus on continued action in time. This may be illustrated by expressions such as *an upright stance* and *the Earth revolving around the Sun*, respectively. As it may be, all schemas consist of a very limited number of elements and simple relations (Johnson 1987: 28), and "are flexible […] in that they can take on any number of specific instantiations in varying contexts." (Johnson 1987: 30). Elaborating on that somewhat, it has been stated that an image schema has "an internal logic of its own that makes it analyzable […] and extensible" (Stadler 2020: 145). Such a capacity of schemas to extend their scope suggests that they can enter specific complex relations with other schemas, an idea that the present approach aspires to develop, based on Cienki (1997) and Cameron & Stelma (2004), among others. In a recent paper, these interactions were observed from the point of view of type, where the most prominent of three options appears to be the one labeled as *Merge*, which entails combining atomic entities into complex image schemas (Hedblom et al 2019: 283–284). Similarly, image schemas may be said to combine into 'profiles' (Gibbs 1994; Oakley 2007), combinations such as RESISTANCE-CENTER/PERIPHERY-LINKAGE-BALANCE-VERTICALITY. Our approach shall try to support the idea that image schema families like these work across cognitive modalities, as well.

While the thesis that schemas interact, and dynamically, is very common, the incorporation of the factor of scalarity in their interaction is not often found in the literature. Rather, SCALE is most often seen as a schema like any other, "ubiquitous" (Clausner & Croft 1999: 22), "a variant of the PATH schema" (Clausner & Croft, 1999: 17) or a "superschema" (Grady 2005: 48). The key difference between our approach and that of Grady (2005: 44) is the fact that he perceives "image schemas […] as mental representations of *fundamental units of sensory experience*", whereas SCALE can be perceived as something transcending the merely sensory and combining the sensory experience with the cognitive (non-sensory) processing even at the early stages of mental imagery formation. This, among other reasons, is what classifies the image schema SCALE outside the known groups. Moreover, Grady considers scalarity as a property necessarily shared between two domains in the metaphorization process, addressing these properties as "superschemes" (2005:47), and concluding rightfully that the "…shared superschematic structure does not constitute the 'similarity' on which metaphors are based."

However, a very important aspect of SCALE that presents a difference between Grady's approach and this one is that for us it refers to a dimension which crosscuts many different types of perceptual and non-perceptual experiences. Thus, it may indeed be considered a parameter, capable of developing interactive relations with schemas and specifying the intensity of image-schematic complexes. In the present approach, therefore, *scale* is not a schema proper, but rather a higher-order cognitive principle involved in the control of intensity of other schemas in the expression. Thus, we hypothesize that the conceptualized FORCE can most likely range from very weak, as in a *tickle,* medium, in *pressure*, and high, e.g., a *crush*. Likewise, for many schemas scalarity also inheres the notion of *valence*. On such a view, one end of the scale implies a negative value, the other a positive one, and a neutral, 'zero' value is postulated exactly in the middle. Thus, a FORCE can be negative (in music: pianissimo), positive (fortissimo) and neutral (an unmarked, 'normal' loudness range), again causing subtle affect-related effects during perception: from the gentleness of a soft passage to the often harsh and piercing impact of very loud musical excerpts; from the positive and jocular connotation of *tickling* to the brute force of *crushing*, which includes its metaphorical extensions, as in political discourse (*crushing a* *rebellion*). A view of conceptual valence similar to this one has been offered in linguistics before (Krzeszowski 1993: 310), and shall help us determine the affective nuances caused by the activity of such scalarized schemas in interaction.

In the domain of psychology, "affect plays a central role in human experience" (Gray & Watson 2007: 171). According to *the circumplex model of affect* (Russell 1980), one of the dominant dimensional models, affect consists of two main components: (i) *valence,* which marks the distinctions on a scale pleasure-displeasure, and (ii) *arousal*, which is defined on a scale from activation to sleep. Consequently, the resulting affective experience is the product of the combination of these two dimensions. Based on *the PANAS model* (Watson, Clark & Tellegen 1988), which is another dimensional model of affect, a distinction is made between the PA (positive affect), understood as a combination of pleasure and arousal, and NA (negative affect), which reflects a combination of displeasure and arousal (Russell, Weiss & Mendelsohn 1989: 498). Russell (2003) and Barret (2006) discuss *the core affect model*, one of the constructionist models of affect, according to which the construction of emotional experience is licensed by the categorization of core affect. Russell (2003: 147) provided a definition of core affect that is very close to the treatment it had in the dimensional models, in that it represents "a neurological state that is consciously accessible as a simple, nonreflective feeling that is an integral blend of hedonic (pleasure-displeasure) and arousal (sleepy-activated) values" (Russell 2003: 147).

Overall, the affective experience appears to be guided by the two main dimensions (i.e., valence and arousal), and the categorization of such basic experience gives way to the construction of emotional content. In the context of the present proposal, we include the higher order cognitive principle of *scalarity* which is hypothesized to amplify or diminish (depending on the actual context) the components of the overall affective experience in the process of meaning construction. This, in turn, could provide a direct link between affect and image schemas, with scalarity functioning as an 'intermediary' mechanism between the affective and cognitive dimensions involved in the process of meaning construction in various settings (not limited to language). Such line of reasoning is further supported by the fact that "the circuitry that instantiates a core affective state is widely distributed throughout the brain, and includes so-called 'cognitive' areas" (Duncan & Barrett 2007: 1185).

The distinction introduced above, between marked and unmarked scalarity (as in a pianissimo or fortissimo seen against a 'normal', 'neutral' dynamic range in music) brings about another idea that we are proposing: that gradience is not an obligatory element of schema interaction. In other words, a zero value implies that scale is simply switched 'off'. Only if it gets switched 'on', and reaches a level on a gradient, toward either the positive or the negative valence, does additional affective nuance creep into the musical or linguistic interpretation: softest or loudest music, tiniest to firmest pressure, and their corresponding linguistic descriptions. Thus, scalarity involves a degree, but it needs to be called ('turned on') by the cognitive system first in order for this degree to manifest at all. We find support for such an additional theoretical extension – according to which scales are optional – in the relatively late emergence of scalarity in child development. Namely, SCALE as image schema is not to be postulated in children until later in childhood (Mandler & Cánovas 2014: 8), nor is it necessarily innate, at least in language acquisition (Mandler 2005: 139–146). Only subsequently in life do humans seem to become 'attuned' to variations in intensity (Johnson 2005: 21).

Thus, we postulate a two-step dynamic interaction. In the first step, it is determined whether the *scale* parameter is set as *on* or *off*. As a rule, in the former case, profiled scalarity fosters an emotional involvement in the interpretation of meaning, roughly corresponding to arousal in the circumplex model. This does not normally occur if scale is inactive. Consider examples from our set of 100 newspaper headlines taken from the *Guardian* and the *New York Times* between 27 May and 1 June 2022:

1. a. *Inflation hits its highest level*

   b. *Israel signs Trade Deal with UAE*

In the first case, (1) a., the FORCE and PATH schemas interact: one needs to initiate the motion that propels inflation upward; and then the upward path is created by the realization of this movement. Importantly, both components gradually heighten here – the FORCE needs to increase as much as possible to overcome gravity, and the PATH has to stretch far out too, until both at one point reach the summit, i.e., the endmost point of scalarity. This conceptualized strain serves as an additional factor to complement the mere denotational interpretation of meaning. It results in the inference in the parser that inflation is causing a serious problem, which may in turn induce unpleasant sentiments. So, the expression does not only metaphorize increase in prices as a physical movement upward; it also calls an affective layer of interpretation, facilitated by the profiled scalarity of the image schemas it inheres. In the second statement, (1) b., the signing of the trade deal is merely reported, where any schemas (presumably LINK) underlying a metonymic interpretation (it is not *countries* that sign the deal, but their human representatives) do not appear to assume a scalar value. The affective impact of such a statement, at least when devoid of context, should be negligible.

When and if *scale* is switched *on*, which is the second step in the interaction, it makes image schemas it interacts with gradable, providing the elements that are needed for the conceptualization, e.g., an increasing or decreasing force, progression from a starting point to an end point, or gradual movement from the outside point relative to a container into the container.

Exemplifications of the former two could be found in the headlines from newspapers as in (2) a. and b., referring to increasing and decreasing FORCE, whereas c. is the manifestation of the CONTAINMENT schema in correlation with *scale*.

2. a. *Prices Hit New Highs* / *U.S. Retakes Top Spot in Supercomputer Race*
   b. *U.N. Human Rights Chief Tempers Criticism at End of China Trip*
   c. *Even as Russia Bears Down in the East, Some Ukrainians Stay Behind*

The statements in (2) a. reinstate the situation with inflation above. Prices hitting high levels will likely cause the same negative sentiment as does soaring inflation, whereas the retaking of a top spot by the U.S. may cause a very pleasant or a very unpleasant emotional reaction – depending on whether the reader comes from the States or, say, China. In (2) b., the "tempering" of criticism effectively reduces the force of expostulation, hopefully making the situation for the Chief more pleasant than one would originally have expected. Finally, in (3) c. the PATH and FORCE of Russia's attack, contained to the area of 'the East', reach a high scale ('bear down' – 'proceed forcefully toward something', a 17th century nautical use), again causing a pleasant or unpleasant reaction depending on the side in the conflict taken or supported by the reader.

In the examples above, we have seen that schemas were originally proposed as a contribution in linguistics. Yet they also work across cognitive modes. In language, schemas normally require an interpretive component (e.g., 'What does *bear down* mean?'), complicating the case for a formal analysis. In music, they can be inferred from the structure of the stimulus alone, which simplifies their potential formal and computational description. Some of the currently available insights on conceptual schematicity in music include Antović (2009), in which a series of schemas, notably DISTANCE, PATH, and FORCE are inferred from Serbian and

Roma musically trained and untrained children's verbal descriptions of musical opposites, such as five-note soft and loud segments. The procedure is replicated in Antović, Bennett & Turner (2013), which shows that north American sighted and congenitally blind children detect a schematic basis in musical segments equally well, usually resorting to the perception of spatial relations. Brower (2000) uses Johnson's image schema theory as partial inspiration for her all-out approach to musical meaning, yet expands on it by introducing a distinction between *musical* schemas, pertaining to intramusical hierarchies, and *image* schemas, cross-modally connecting musical and bodily experiences. For instance, she presents triadic, diatonic, and chromatic relations within a scale as belonging to three concentric circles of stability, or schematic NESTED CONTAINERS. These in turn motivate the CENTER/PERIPHERY image schema, exerting varying levels of stability (and thus the FORCE of attraction towards the center). Moving from the first of these schematic constructs to the third, the listener appears to progress from a purely intramusical to a more palpably extramusical, embodied experience. Cox (2016) takes the embodiment thesis a step further when he proposes an EXERTION schema to relate muscular effort and musical movement, ultimately subsuming all music cognition. Somewhat similarly, Larson (2012) bases an entire theory of musical meaning on the notion of FORCE (namely gravity, magnetism and inertia), influenced by Talmy's (1988) program in cognitive semantics, yet he equally pays attention to spatial schemas, as when he discusses the metaphorical movement of time against the observer and its extension to musical movement. This idea is based on his earlier work on musical motion in music and text in a Beatles song (Johnson & Larson 2003). Saslaw (1996) suggests that most instances of modulation in music are grounded in the SOURCE-PATH-GOAL SCHEMA, while CONTAINER may also be present in deciding "which way" a modulation will go, given the various cognitive pressures to resolve the harmonic tensions in one direction or the other. Finally, in his thorough work connecting cognitive linguistics and music theory, Lawrence Zbikowski has paid image schemas due attention as well. He was among the first to point to the importance of VERTICALITY as an abstract structure motivating the common option for conceptualizing musical pitch as height (Zbikowski 2002: 69). He equally explored how schematicism inhered in a series of musical elements, including pizzicato, may be suggestive of the act of knocking in an instance of text painting in Bach (Zbikowski 2008). More recently, he has upgraded the approach based on Mark Johnson's schemas to that inspired by Lawrence Barsalou's perceptual symbol systems theory (Barsalou 2004). On that view, "sonic analogs of music can be seen as a form of conceptual knowledge that is derived from perceptual information, and that is connected to other kinds of conceptual knowledge through shared configurations of properties and relations" (Zbikowski 2017: 37). This author's important example is Sagreras' *El Colibri*, in which elements such as rapid repeated notes and sudden shifts of pitch along the strings evoke the sound and movement typically produced by a hummingbird.

It appears, then, that using image schemas in music theory and cognition has provided many fruitful insights. In what follows, we propose a novel approach to schemas in music and language to address the question of affective meaning construction.

3. Dynamic approach to meaning generation in music and language

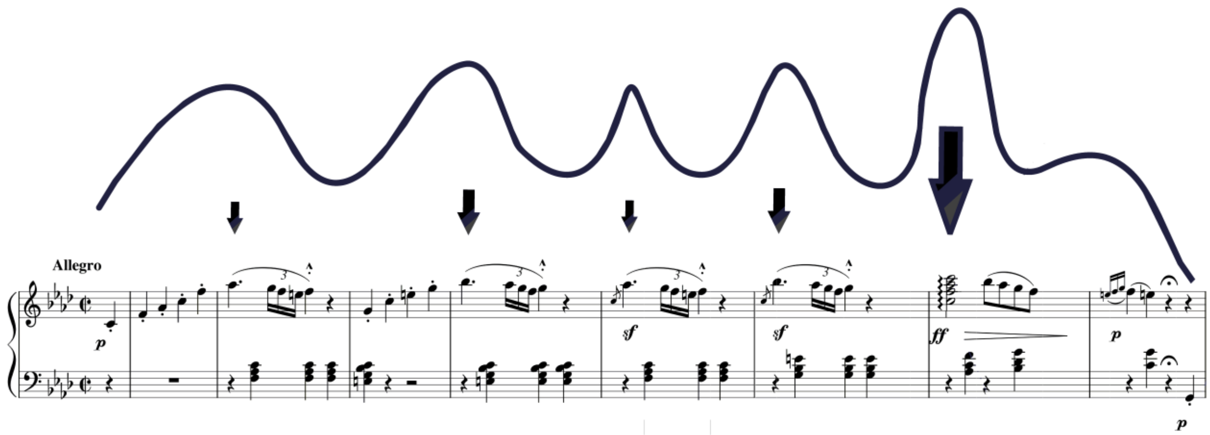
A lot of classic research in cognitive linguistics notes that the nature of schemas is dynamic (Johnson 1987; Clausner & Croft 1999; Stadler 2020). After all, starting at a "source" and traversing a "path" to reach a "goal" obviously requires some continual, structured effort evolving in time. Yet, this aspect of schematicity is rarely profiled in attempts at formal analysis, likely because a continual dynamic is very difficult to capture by any discrete system. For instance, in the expression *beating around the bush*, the interaction of FORCE, (circular) PATH and CYCLICITY progresses simultaneously, so that one cannot easily differentiate between these schematic notions at any given point in time. Yet the three schemas can still be *named*, *conceptualized*, and *represented* as discrete constructs. To address some potential solutions for a formal approach to multimodal semantics emerging from this duality, we start with music cognition, as musical schemas are inferable already from structural analysis. Then we progress to a corresponding example from linguistic semantics to illustrate how image-schematic interaction may be formalized across modes.

Our excerpt of choice is Beethoven's Piano Sonata No. 1 in F minor, largely because it is well known to professional and lay audiences and because it represents a common example used to explain what a musical phrase is by theorists from the Western world. Obviously, such a choice suggests that our approach is confined to classical occidental homophonic music. A broader cross-cultural account and analysis of different musical textures (e.g. polyphony, as discussed in Powers (1980) with regard to a comparable approach by N. Ruwet) remain an inspiration for future work.

Consider the sonata's opening (Figure 1, adapted from Antović 2022: 45).

Figure 1.

The Opening of Beethoven's Piano Sonata No.1 in F Minor with a Notation of Real-Time Changes in Perceived Musical Energy above the Score



There is minimal consensus in the complex debate on musical meaning that the first glimpse of signification emerges from changes in musical tension which the parser (idealized listener) experiences while perceiving a piece (Bernstein 1976; Jackendoff & Lerdahl 2006). In this particular case, for a number of reasons rooted in musical structure (roughly the inference of dominant metrical, melodic, and harmonic factors followed by changes in musical loudness and articulation, see, e.g., a Generative Theory of Tonal Music, GTTM, Lerdahl & Jackendoff 1983), the parser is expected to experience four instances of gradual rise of tension towards peaks of segments lasting one or two measures (represented as small arrows above the notation system), culminating in the principal peak in the seventh measure, on the arpeggiated F-minor chord given as a fortissimo half-note (pointed at by the largest, fifth arrow). This perceptual high point may be said to exude the strongest amount of musical "energy". "Energetics" is a relatively recent concept in music theory, traceable perhaps to early 20th century German writers (Rothfarb 2002). Jackendoff and Lerdahl quote as inspiration a prominent such author, Heinrich Schenker, but also classical Gestaltists such as Wertheimer, Koffka or Köhler. Koffka (1936: 180) writes that "if greater articulation implies that more energy is consumed in the process, then these greater forces must have liberated more energy, just as an electromotor doing work against forces uses up more energy than an idling motor.“ The consequence for music theory is that points of pronounced articulation may liberate more energy as well (note that articulation in Koffka’s sense means the physical effort in producing a musical sound – this could be actual articulation employed in music theory such as staccato, but also other kinds of effort such as loud dynamic, pitch rise, or dissonance). Closer to the cognitive linguistic epistemology, similar positions are found in Larson (2012), who equally finds a connection with energy in terms of musical forces and leaps, and Cox (2016), who relates musical schematic energy more directly to muscular movement. There is a further complication with Jackendoff and Lerdahl, however. The highest of the four abstract musical structures of interest in their work, “prolongational reduction”, does not deal with individual note properties, but rather their *interrelations*, describing listeners’ intuitions about the constant sequencing of tension/relaxation and continuity/progression in the musical flow. They do not explicitly use the term “energy”, but write that prolongational regions give off “power” (e.g., Lerdahl & Jackendoff 1983: 272). We elaborate on this proposal in the following section.

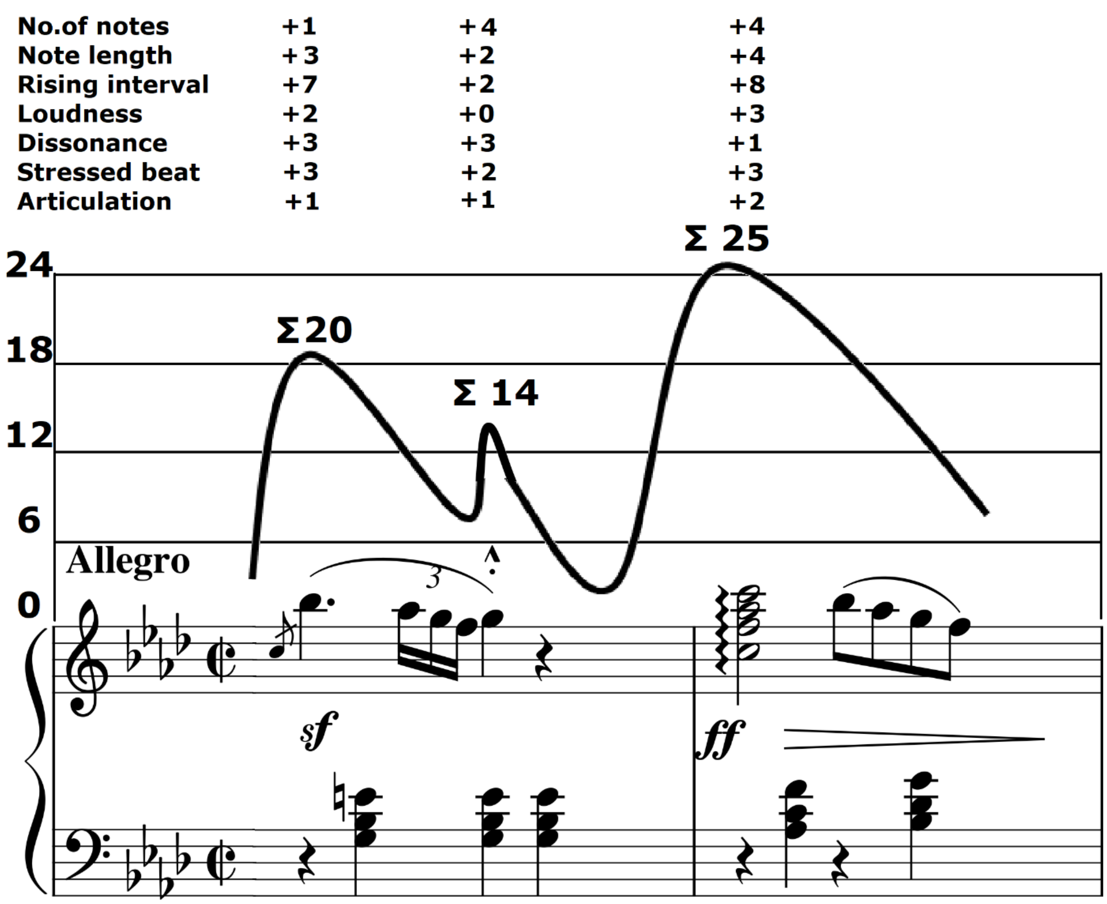
As it may be, after the peak is experienced in measure 7, our proposal above suggests that the energy quickly decreases in measure eight, just before the theme gets repeated in C minor in measures nine and on (not given in the present excerpt). The dynamic variability of musical affect ensues from gradual increase and decrease of perceived musical energy between these points of strongest articulation, and may be represented as a curved line (above the arrows and the notation system).

One may compute numerical values denoting the amount of perceived musical energy by focusing on structural elements present in the notation that require more or less effort in the production of sound, e.g., ascending or descending intervals, stressed or unstressed beats, articulation symbols such as staccato and legato, loudness markers, such as fortissimo and piano, or harmonic dissonance and consonance. The idea is that musical elements employed in scores may be broadly divided into *marked* (salient, stressed), boasting potent musical energy, and *unmarked* (lax, unstressed), suggesting that the energy at those points is low.

Consider again just the sixth and seventh measures from the Beethoven example (Figure 2).

Figure 2.

The Sums of Musical Elements Exhibiting "High" Musical Energy in the 6th and 7th Measures of the Beethoven Sonata Opening



We define seven factors for computing musical energy in this segment: the number of notes played simultaneously (one point per note); the duration of the targeted note(s) – here we assign one point for an eighth note, which we take as occupying the mid position in the note length continuum; above it we find quarters (+2), halves (+4) and wholes (+8) and below it sixteenths (0.5), thirty-seconds (0.25) and sixty-fourths (0.125) – the numbers stand for length ratios; longer and shorter notes than these are quite rare in the literature); the size of the ascending musical interval (one point per degree); the stressed location of the beat (one point for unstressed, two for relatively stressed and three for fully stressed); loudness (e.g. +1 forte, +2 sforzando, +3 fortissimo); the dissonance of the scale degree realized by the targeted note(s) (+1 for the resolving tonic, +2 for the subdominant, +3 for the dominant); articulation in the narrower sense (e.g. staccato or arpeggio are +1, if the articulation is realized on more than one note playing simultaneously, the value is +2).

Let us focus on the three note positions resulting from pitch elevation in the segment. The initial B flat in the sixth measure comes from the following marked factors: it is a single note (+1), a dotted quarter-note (+3), it comes as a result of an ascending seventh (+7), it is played sforzando (+2), realizes a dominant (+3), falls on a stressed beat (+3), and is a crushed note (acciaccatura, +1): the total energy value is 20. The second musical sound resulting from a frequency increase is a chord peaking on the marcato G. It is a tetrad (a four-note chord, thus +4), one quarter in length (+2), comes as a consequence of an ascending second (+2), is neutral in terms of loudness (0), realizes a dominant (+3), falls on a relatively stressed beat (+2), and is accentuated as a marcato in the highest note (+1): the total energy value is 14. Finally, the arpeggiated F-minor chord peaking on the fortissimo high C in measure seven occurs as a tetrad (+4), lasts for one half note (+4), realizes an octave jump upward (+8), is played fortissimo (+3), is consonant after a dissonance (tonic occurring after the dominant, +1), occurs on a stressed beat (+3), and is articulated as a broken chord (+2): the total energy value here is 25. One should add that the pause preceding this chord adds to the strength of the perceptual effect as well: the contrast between the silence and the sound enhances the sensation of added energy. Alternatively, one may contend that there is no literal pause in the sequence, as the arpeggiated chord in the treble clef connects to the dominant chord in the bass. In that case, however, the energy additionally rises as well, only due to another reason: the frequency jump from the highest note of the dominant chord in measure 6 (E_4) to the highest note in the broken tonic in measure 7 (C_6). Therefore, depending on the interpretation, the increase of energy here could amount to as much as +13 instead of +8.

Naturally, the formalization given here is simplistic, as individual factors can be differently weighted in different musical contexts. For instance, a long note or an arpeggio do not need to assume a marked value (e.g., if the entire piece consists largely of standing notes or broken chords). Likewise, a change in articulation (e.g., a sudden staccato in a fully legato section) may be much more salient than physical force (e.g., if the section is to be played loudly anyway). In such a case the staccato would likely need to receive more ‘points’ than the forte. Yet the important fact remains that the interplay of factors provides an opportunity for a formal account and GTTM remains a very good system to base such an effort on.

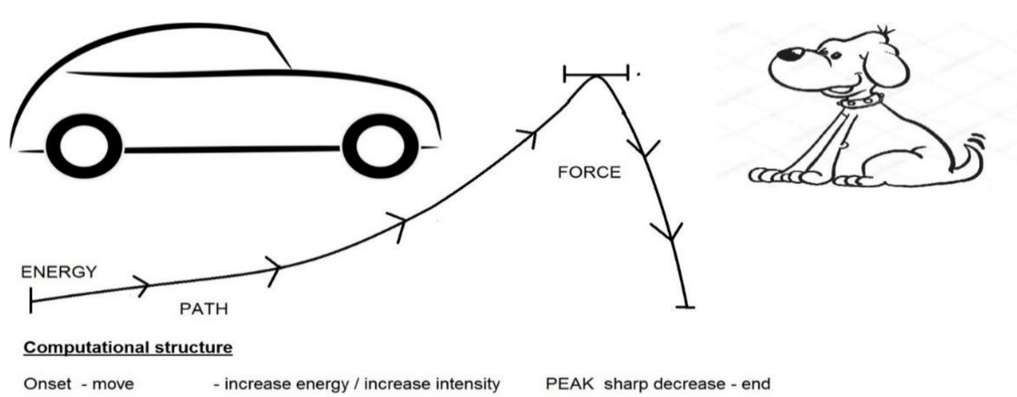
In language the situation with formalizing online meaning-generation factors is complicated by the fact that schematicity operates as part of referential (predominantly lexical) semantics much more than of grammatical (predominantly sentential) structure. Thus, unlike music, whose meaning derives from form, language boasts *double articulation*: in it, the structural elements combine according to some formalizable principles, yet it is almost never the case that semiosis emerges *directly* from the employment of those elements. For instance, the fact that *bush* consists of three ordered speech sounds of English has little or nothing to do with the item's denotation ('a shrub with stems of moderate length'). Rather, the sound-to-meaning connection originates as arbitrary and is ultimately a matter of social consensus.

The consequence is that, in language, one may not easily infer schemas from structural relations, but must rather consider referential meanings present in the mental lexicon. This makes the process less tractable, dependent on native speaker's intuitions, and often impressionistic. Yet such a complication does not preclude the fact that in linguistic semantics one may also find dynamic schematic relations similar to the "changes of musical energy" from the Beethoven example above.

Consider the following expression: *A car sped up and hit a dog in the street*. On the condition that the car stops after the collision, the situation may be imagistically represented as in Figure 3.

Figure 3.

A Graphic Representation of Levels of Conceptualized Energy in the Car/Dog Event

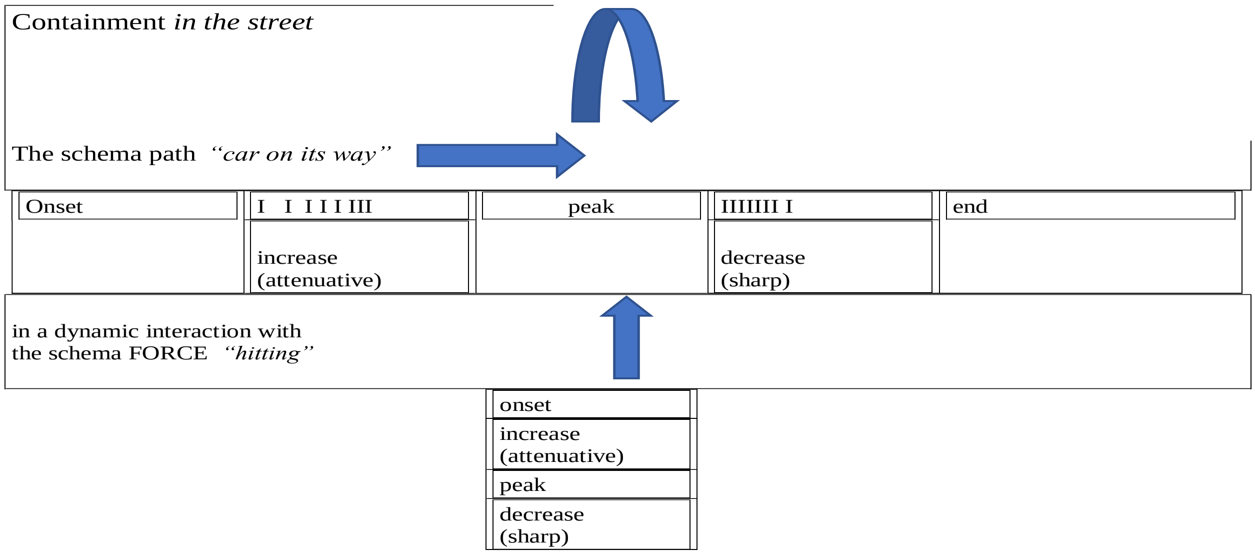


The scenario involves an object moving along a trajectory, increasing its energy to enhance the speed of movement, at one point reaching the peak of this energy, and finally colliding with a static object, after which the movement quickly collapses to a standstill. If one focuses just on the online change in the amounts of energy, one may represent this activity as another 'semantic curve,' quite similar to that discussed with regard to the 7th measure in the Beethoven example.[1] Accordingly, the continual process may be further represented as a schematic interaction of PATH (the trajectory along which the speed of the car increases) and FORCE (producing the movement in the first place, and then increasing, to peak at the moment of collision). Likewise, the path in this case too goes both, conventionally, "rightward", embodying the car's movement toward the dog, and, metaphorically, "upward", suggesting the gradual energy increase as the car speeds up. An added factor here is the schema of CONTAINMENT, as the activity takes place in the enclosed space of a street. A possible tabular representation would look as in Figure 4.

Figure 4.

A Possible Semantic Interpretation of the Sentence "The car sped up and hit a dog in the street."

[1] In fact, there might be even more multimodal aspects to this scenario, as during the acceleration process the sound produced by the car would become gradually louder and higher. This would in turn serve as onomatopoeic signification, fortifying the connection with music. Naturally, such a possibility would remain part of the interlocutors' possible mental simulation of the scene. It is not contained in the language structure per se.



Similar to the arrows denoting perceptual peaks in the musical example above, this graph provides a rightward arrow as an indicator of movement, an upward arrow as the landmark of collision, and a semicircular arrow as the token of enclosure ("in the street"). Likewise, the graph attempts to capture individual moments at which the speed of the car gradually increases as progressively reduced distances between "I" symbols just below the rightward arrow, in the "increase (attenuative)" box. In a similar fashion, after the peak speed is reached, the "decrease (sharp)" box presents a series of equidistant "I" symbols in the closest position to one another – again suggesting the highest, and stable speed, followed by a single, final, separated "I" that represents the sudden reduction of velocity, leading to a final halt (the "end" box).

Ideally, one would want an AI routine which would capture the key moments in this series of events and calculate the sums of marked and unmarked elements, as in the musical example above. However, the problem is that there are no formal factors which could tell us anything about the dynamic structure of the schematic complex here. How much is the car speeding up? What is the "level of the force" that needs to be produced in order for the dog to be hit? What “threshold“ does this force need to reach so that the dog would lose its balance? And so on. There is nothing in the auditory signal or the written sentence that provides such information. So, the analysis based on an automatic detection of formal factors alone cannot work.

A possible way out of the problem may be found if one should move away from seeking the formal factors *in the signal* (or written language), but rather focus more deeply on *inferred underlying structure*, in which individual schemas are combined with the higher-order notion of scalarity. This is the topic of the following section.

4. **Improved Analysis: Image Schematic Complexes and Scalarity**

In this part, we explore the idea that image schema combinations, enhanced by the higher-order parameter of scalarity, may be useful both for tracking real-time changes in meaning and for inferring the basic emotional load sparked by musical structures or linguistic constructions. In terms of combinatoriality, it is quite obvious that any more complex events could never be interpreted on account of a single or a pair of image schemas engaged in the conceptualization, but rather by a group of those interacting in a specific way, following certain common patterns (Hedblom et al 2015). As for scalarity, as we have seen, cognitive linguists and psychologists have often considered SCALE a schema like any other. This has occasionally caused theoretical inconsistencies. For instance, the idea that SCALE is just a modified PATH schema (Johnson 1987) would cause serious problems in interpreting the sentence *The more he climbed the ladder the faster he moved*. In this scene, the movement upward along the path is just one component of the story; the gradually added force increasing the agent’s speed along the way is a separate, and scalar, factor. Conflating the two would result in the notion that both progressing upward and increasing speed represent changes along a scale. Obviously, such an interpretation fails to capture the semantic complexity of the expression – the conceptual *movement* realized by the employment of the PATH schema and the *modification* of this movement embodied by the notion of the SCALE.

A possible way out of such inconsistencies, in our view, is that scalarity should be given a higher-order status, to work as a parameter modifying other classical schematic constructs regularly employed in cognitive linguistics. Thus, added *scale* may increase the FORCE of movement, but also the length of PATH, nearness of LINK, stability of BALANCE, or scope of CONTAINMENT, to name but a few possibilities.

Let us now consider how schematic complexes and scalarity may work in the Beethoven musical example we have used above. To reiterate, one way to look at the changes in musical energy is to focus on sums of intramusical factors at individual notes and derive the intramusical semantic curve from them, as described in the previous section. A somewhat changed perspective is reached if one focuses on *combinations* of musical notes and image-schemas emerging from them. There, the analysis does not track individual musical elements, but rather their interrelations. As we shall see, this allows for a more nuanced appreciation, not the least because it breaks down the monolith sensation of musical energy into a set of ontologically diversified – and thus proto-conceptual – constructs. For the purposes of the present analysis, we focus on five schemas mentioned above: FORCE, PATH, LINK, BALANCE, and CONTAINMENT. [2]

[2] One may wonder here how a program on image schemas in music, inspired by Lakoffian cognitive linguistics, can build on the insights of the *generative* theory of tonal music. The fact is that instead of "lower level" generativist constructs of GTTM (grouping, metrical and time-span structure), we start from this theory's final construct of "prolongational reduction", which aims to capture intuitions of tension and relaxation, continuity and progression in the musical flow. The construct goes beyond the pure grammatical parsing and predicts elementary affective responses to music by the listener. This in turn broadens generativist formalism, to encompass the phenomena of musical emotion and meaning. With conceptual schematicity, which we propose as the next step in this effort, the connection between the two apparently contrasted research programs becomes obvious. While cognitive linguists claim image schemas "map spatial structure onto conceptual structure" (Oakley 2007: 215), Jackendoff agrees

So how can interlinked and scalarized schemas help in the quest for what motivates affective meaning in music? While the focus on "energy" or even affect alone is constrained to what authors have labeled "formal" (Bernstein 1976) or "intramusical" semiosis (Koelsch 2012) - the perception of ineffable changes in the musical flow - the introduction of schemas allows one to also consider cross-modal interactions between music and extramusical phenomena ("iconic" signification, Bierwisch 2009). Typically, the progression of musical elements in time invokes the schemas of FORCE initiating the musical events and PATH emerging from the sequencing of such events, which together motivate the more complex metaphor of musical movement (Johnson & Larson 2003; Antović 2018). The interaction of the two schemas is dynamic since, for instance, in order to reach the point of the highest force in the seventh measure in the Beethoven, the music needs to surmount a considerable distance, gradually overcoming paths that, in standard Western musical notation and convention at least, go both "rightward" (in time) and "upward" (in frequency). Of course, other cross-cultural options are quite possible (e.g. different directionality of the movement or some sort of shape transformation, Antović, Mitić & Benecasa 2020), yet we do not deal with them in the present work. The question for formal analysis is how to capture these intuitions through discrete values, such that could be used to predict changes in both affect and cross-modal imagery in the perception of other musical pieces as well. This is required, for instance, to design experiments with participants or to program AI routines.

"the concept" is largely spatial, adding that it "must be some sort of finite schema that can be compared with the mental representations of arbitrary new objects to produce a judgment of conformance or nonconformance" (Jackendoff 1990:9). Several epistemologal disagreements, of course, remain: "cognitive linguistics tends to view cross-field parallelisms as derivational. [...] By contrast, I view them as parallel instantiations of a more abstract schema" (Jackendoff 2002: 360). This entails that Jackendoff considers schemas amodal and potentially universal, while cognitive linguists typically see them as experiential and culture-dependent. In related experiments, a strongly relativist view is defended in Dolscheid et al. (2013), a balanced position in Athanasopoulos & Antović (2018), and a fully Jackendoffian account in Antović, Mitić & Benecasa (2020). Yet, these subtleties irrespective, one can claim that an approach to musical affective meaning based on conceptual schemas may be equally attractive to proponents of conceptual and cognitive semantics.

What we need here is an analysis that will go beyond the mere calculation of energy change and that will incorporate finer-grained factors of musical experience. In line with our two-step approach described in section 2, let us assume that the schemas can take two valences (the schema operates “above” or “below” its zero point) and three individual levels of intensity (*scale*) for each valence, notated by a series of up to three plus and minus signs in a row. The two valences are inspired by many phenomena in music theory treated as semantic antonyms (e.g. piano-forte) and the three-level gradation comes from both the conventional notation of loudness in a musical context and the corresponding levels of adjective comparison in many languages (the two valences can thus be interpreted as binary/parametric and the three scalarity levels as more gradable, thus non-parametric, cf. Ruwet & Everest 1987: 16). An important difference from the analysis in the previous section is that these schematic factors are based on tonal *relations* and not absolute values of factors in individual notes. Thus, our FORCE will pertain to levels of *change* in the dynamic: *F*--- = loudness reduction by three levels or more (e.g. unmarked dynamics to piano pianissimo), *F*-- = reduction by two levels, *F*- = reduction by one level, *F0* = no change, *F+* = volume increase by one level (e.g. pianissimo to piano), *F++* = increase by two levels (e.g. unmarked to fortissimo), *F+++* = increase by three levels or more.

In an analog fashion, PATH implies musical movement through a vertical trajectory and may be formalized depending on the size of individual intervals (steps) that are involved in the melodic segments: if we take an octave to be a default vertical movement range, and keep the classification into three levels, we notate *P0* if the note is merely repeated, *P+* if the step covers an interval of up to a fifth, *P++* if it extends up to an octave, and *P+++* if it jumps by an octave or more. Paths cannot normally assume a negative valence, but their direction may be further specified, e.g., in music, as going *up* or *down* which we may notate with an additional specifier, as needed (e.g. P+ spec: *down*).[3] LINK in music involves the change in the degree of articulation in which the musical notes are (not) connected to one another, along the following spectrum: marcato, staccato, portato, no articulation provided (unmarked), legato marked by a slur, legato marked by a slur and at least one more musical symbol (e.g. the pedal sign). *L*--- = change by three degrees or more towards the "disconnected" articulation (e.g. legato into staccato), *L*-- = a corresponding change of two degrees, *L*- = a change of one degree (e.g. portato into staccato), *L0* = no change in articulation, *L+* = a change from a less to a more "connected" articulation of one degree (e.g. an unslurred unmarked articulation into a legato with a slur), *L++* = a two level such change, and *L+++* = such a change of three or more levels.

Musical BALANCE depends on harmonic relations, yet represents a slightly more complex phenomenon. Here we follow Lerdahl and Jackendoff's view that prolongational reduction (responsible for musical affect) is based on two factors: tension/relaxation and continuity/progression. The former construct deals with chordal stability, where dissonant chords tend to be more tense. The latter focuses on chord connections, where any change of the harmonic root of two chords means a progression. Incorporating these two factors, we define the following situations, progressing from decreasing toward increasing stability: *B*--- = two dissonant chords are linked, where one has at least one out-of-key note, e.g. a dominant and a secondary dominant, *B*-- = two dissonant chords are linked, remaining in the key, e.g. a subdominant and dominant; *B*- = a consonant and a dissonant chord are linked, e.g. the tonic and the dominant; *B0* = continuity, no change of the chord root; *B+* = a dissonant and consonant chord are linked, e.g. the dominant and tonic; *B++* = two consonant chords are linked toward the

[3] Rightward – thus horizontal – paths are, of course, ubiquitous in music, too: in the Western conceptual system and notation at least, all music moves "to the right" all the time. We do not focus on that phenomenon in the present paper, though.

less stable chord, e.g. the major tonic and its parallel minor; *B+++* = two consonant chords are linked toward the more stable chord, e.g. the parallel minor and its major tonic.

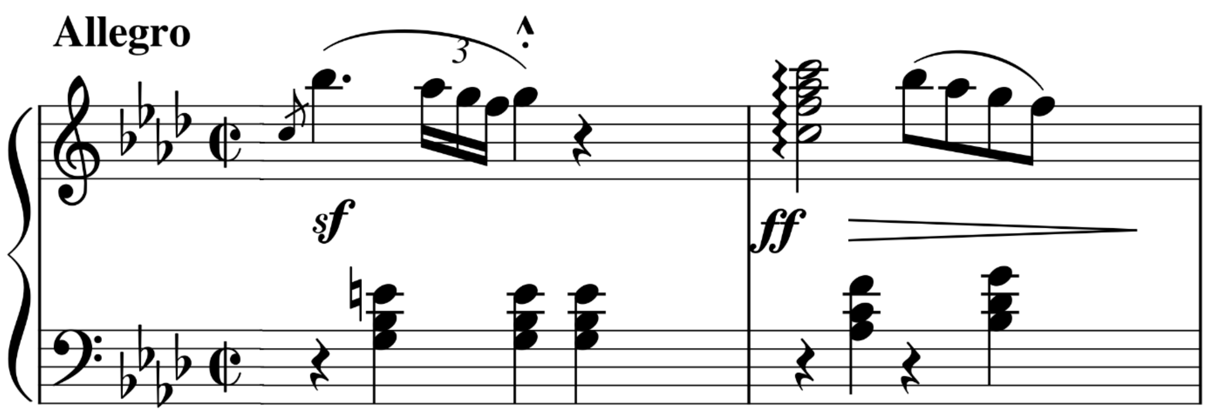
Finally, CONTAINMENT can also take only a positive value. In our musical analysis, we use it for specific instances of short, ornamental musical notes immediately announcing the following main note, which are thus in some sense "contained" by these longer notes. Passing tones are a typical example of this phenomenon. Depending on the number of notes in the anacrusis resolving into/belonging to the final note in such clusters, containment can also range from *C0* = no containment, over *C+* = a single note announcing the final note, e.g. an appoggiatura, to *C++* = a dyad belonging to the final note, e.g. a very short third and fifth degree double-note resolving into the upper tonic, to *C+++* = three or more notes working as constituents of/ announcing the final note e.g., a triplet or an arpeggio.

An analysis utilizing just these five schemas and the notion of scalarity, on the example of the 6th and 7th bars from the Beethoven sonata, would look as in Figure 5. We again focus on the three "peaks" of interest from the previous section.

Figure 5.

The Interaction of Five Scalarized Image Schemas in the 6th and 7th Bars from Beethoven's Piano Sonata No. 1

Path	P++	P+	P+++
Force	F+	F0	F++
Balance	B-	B0	B+
Link	L0	L--	L+
Containment	C+	C+++	C+++



The graphic presentation may remind one of structuralist models (e.g., Ruwet & Everest 1987) in that we move away from actual notated data to postulated and/or inferred make-up of mental representation. Note, however that complex hierarchies inferred from the score are not our direct interest in the present work. Rather, each of the lines above the musical notation follows the development of one schema (PATH, FORCE, BALANCE, LINK, CONTAINMENT), where sequences of plus and minus signs show points at which the change of a parameter occurs and therefore particular schemas get more or less strongly profiled (zeros, on the other hand, present situations in which musical parameters do not change and schemas do not operate). The dynamic experience of variable affect in musical meaning construction is now a consequence of three overlapping, yet formally tractable, discretized factors: individual schema valence, individual schema scalarity, and simultaneous activity of up to five schemas at various points in the musical flow (this last factor enables us to calculate the 'overall', total scalarity score as well).

In the context of the dimensional models of affect, scalarity is expected to modify both valence and arousal, thereby diversifying the overall affective content of the construed musical meaning. Hence an analysis based on scalarized image schemas aims to go beyond musical energy as a single factor motivating the listener's affective response. In that sense, we now break the construct into two mutually related phenomena: *intensity*, which generally rises with increased total scalarity (and vice versa); and *complexity*, which gets stronger if more schemas are invoked (irrespective of their valence). In such a system, the arpeggiated F minor chord is experienced as the peak of the sonata opening as it has the strongest intensity (+10) and complexity (all five schemas are active). The B note opening measure six is much less intense (+3) though still complex (four active schemas out of five). Finally, the dominant chord with the high marcato G in the second part of measure 6 is equally less intense (+3), but also less complex (three schemas active).

Of course, once again, this analysis is relatively simplistic as it neglects all-prevailing musical recursion: formal structures in music occur on a variety of simultaneous hierarchical levels, so, for instance, whatever is stressed on a "low" tier (e.g., the tactus) does not need to be pronounced at all on a "high" one (e.g., the entire sonata movement). Nevertheless, even if it detects only elementary musical schematization, our analysis still suggests that the more individual schemas are used in short musical segments, and the stronger their total scalarity, the inferred musical affect will be stronger. For another neat advantage, this approach also allows a musical meaning researcher to go beyond mere affect by focusing on individual schemas and the way they may cross-modally map onto bodily or spatial experience. This opens up the way for a *referential interpretation* of the musical meaning that one wishes to get, e.g., from participants in experiments or movie theater audiences listening to soundtracks: a non-legato section gradually increasing by intervals of the second is relatively unmarked (L0, P+). Therefore, if an experimental participant is pressed to verbalize this relation, they might just say that “the music is going up.” On the other hand, a change toward staccato, going up by sevenths, with increasing volume and occasional volume increase is quite marked (L---, P+++, F++). A participant asked to describe “what is going on” in such music would likely provide a much more affectively rich interpretation, perhaps that someone is “ascending vigorously and ominously” (cf. Antović, Stamenković & Figar 2016; Küssner et al 2022). Hence the connection with the main proposed contribution of the present approach: a theory which would use image schemas to touch upon the connotative and affective aspects of meaning generation, where the proposed mechanism of scalarity would directly modulate both the ‘strength’ of the possible range of associative meanings and the degree of the experienced affect. While cross-modal associations in music cognition have been a long-standing topic of experimental interest (from Pratt 1930 to Spence 2020) and while some of this work has been informed by insights of cognitive linguistics generally (e.g. conceptual metaphor, Eitan 2017; Eitan & Timmers 2010) or image-schemas particularly (Antović 2009; Antović, Bennett & Turner 2013), the focus on *discretized* and *scalarized* image schema *combinations* that could partly *predict* the range of possible connotations may provide novel insights in the future.

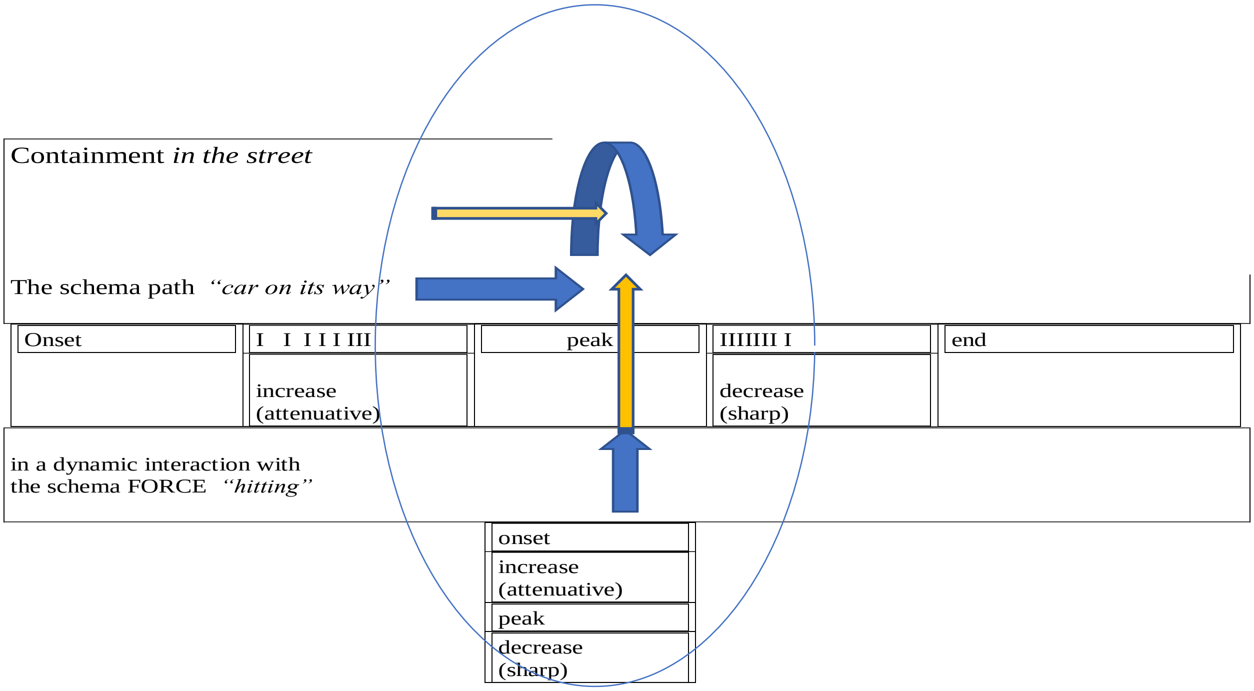
How would this fare in terms of language analysis? We first need to reiterate here that the postulated schemas in our approach primarily pertain to the *event* (i.e. dynamic mental simulation) of whatever it is that is narrated by the referential semantics. Thus we do not analyze the example based on word senses alone, but rather on the holistic interpretation coming from the full parsing of the sentence and attribution of its meaning in the form of mental imagery.

This impinges on our comparison of musical and linguistic semantics. We do not compare structures, as linguistic meaning of course does not (primarily) lie in grammatical structure. Equally, we do not compare combinations of musical elements with word senses. What we do compare is the mental simulation of the event denoted by the linguistic material and the corresponding mental simulation that results in the affective reception of musical structure.

Let us return to the example sentence "The car sped up and hit a dog in the street." As discussed in section 3, the image schematic complex presupposes a stage of concatenation, with PATH and FORCE schemas underlying the movement, while the CONTAINMENT inherent to the phrase "in the street" provides for the background semantic setting of the sentence. Yet the important added factor in the interpretation is the fact that the car speeds up gradually, and that upon the peak of its energy the crash occurs. This may be explained by merging the PATH and FORCE conceptualizations with the parameter of *scale* (represented by the thinner, elongated arrows in Figure 6). To reiterate, unlike in music, in which schematicity emerges already from structure, in the linguistic example it is not the form that is indicative of the scalarity. Rather, the scale must be interpretive, understood from the complexity and the inherent nature of the *event* described linguistically. The entire complex conceptualization is presented schematically in Figure 6.

Figure 6.

Representation of Image Schema Interaction with Additional Elements in the Domain of Scale/ Intensity.



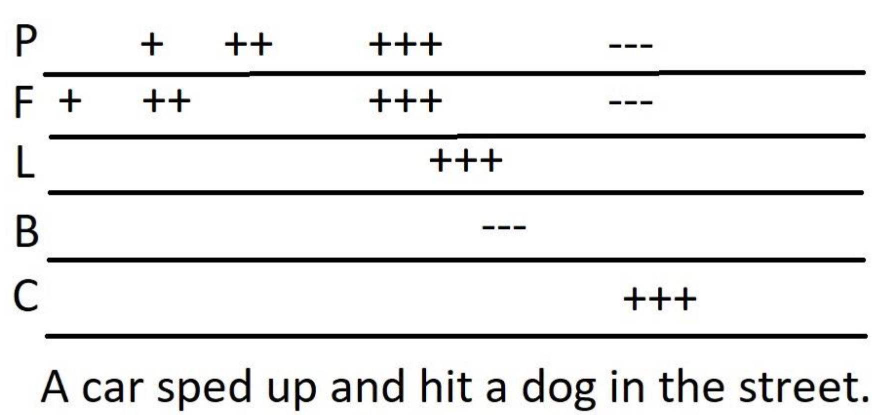
As it can be seen in Figure 6, it is not only the combination of image schemas of PATH, FORCE, and CONTAINMENT that contributes to the formulation of meaning with this complex conceptualization. The element of FORCE demonstrates a variation in terms of intensity which is at its peak at the point of impact of the Trajector (car) with the Landmark (dog). At the same time, the element PATH necessarily entails a progression from the onset point, and through successive points on the way. The Trajector – the moving vehicle – reaches the final limits of its PATH before impact. Finally, the entire interaction of conceptual basic elements occurs within a boundary experienced as a CONTAINER (the street), certified by the preposition *in.* The elongated arrows are an element added to the figure from the previous section, suggesting that both the PATH and the FORCE gradually increase, reaching the final point (collision) at the moment at which the two arrows should touch one another.

However, it appears that here too the image schemas involved need another value which corresponds to their varied manifestation in the conceptualizations. As the schema FORCE is not the same at all the levels or layers of engagement in the processing, but requires a shifted perspective and understanding of a “varied intensity“, the construal may require an additional stage of interpretation. By the same token, in any active conceptualizations of similar kind, the schema PATH is not to be conceived of as a fixed entity in meaning generation , but rather a dynamic entity open to extensions. All this suggests that in order for us to envisage FORCE or PATH as developing, dynamic and discrete values, we have to understand how they merge with the *scale* factor so as to invest scalarity in the conceptual meaning.

A possible analysis analogous to the musical interpretation above would look as in Figure 7.

Figure 7.

A Dynamic Multilinear Representation of the Schematic Structure of an English Sentence



Like in the musical example in Figure 5, the five lines above the sentence present the gradual progression of five image schemas (PATH, FORCE, LINK, BALANCE, and CONTAINMENT), where discrete numbers of pluses and minuses represent discrete points at which the rate of schematicity significantly changes. (To avoid the proliferation of ‘/’ symbols, in this figure, the absence of schemas is simply not notated). What one infers from the figure is that the PATH and

FORCE schemas get gradually strengthened until the moment of collision, where the strongest LINK also occurs (the car physically touching the dog), after which all of the three schemas sharply decrease (including BALANCE, which becomes profiled as [presumably the dog's] lost balance after the collision.) The expression *in the street* provides strong CONTAINMENT as it really encapsulates the entire scene – this has resulted in three plus signs there, as well. Once again, the schemas are meant to coincide with the mental representation of *events*, not the language structure per se; while the three pluses for link are annotated above the actual phrase *hit a dog* in the sentence, one really needs to parse the entire utterance and create a mental simulation in order to understand where the peak of force occurs. Likewise, containment is notated only when the phrase *in the street* is pronounced/read; again, the late positioning of this prepositional phrase does not preclude the fact that the CONTAINMENT schema encapsulates the entire utterance/speech act/visualized scene. Keeping the same logic as in the musical example, the strongest schematic influence, and corresponding affective load, occurs at the point of impact: at this moment, the sum of positive schematic factors is 9; there are three simultaneously operating schemas (FORCE, PATH and LINK) and each of them has the highest positive scalar value (3). In turn, such schema cooccurrence, coupled with the highest scalar values attributed to each individual schema, is also expected to afford a high level of affective content.

Of course, this analysis suffers from drawbacks, as well. While in music the main problem is the recursive nature of the (formal) system, where tense points on one level may be quite lax on another, in language the principal issue is how to extract schemas from combined lexical meanings in all of their contextual complexity. Two solutions come to mind here: one, in which the routine would check definitions in standard dictionaries, where an equivalent higher-order term followed by an appropriate intensifier would be an indicator of the type and intensity of the schema. For instance, 'smash' is stronger than 'break' since it is defined as "*violently break* something to pieces". Thus if 'break' is considered F+, 'smash' would be assigned at least two levels of positive FORCE gradience. As this process can never work perfectly, a separate (or supplementary) method would be the work of human raters on a corpus, such that they could assign schemas to linguistic expressions in context based on a pre-defined set of rules. This is exactly what our team is working on in the present project.

Our procedure looks as follows: define the criteria for assigning schemas to lexical structure as precisely as possible (in pairs of researchers); search for the tokens in a relevant corpus and annotate them; once done, check the solutions with a third researcher; for all instances of disagreement, organize a majority vote. We present examples of how this works from our preliminary selection of 100 recent online media headlines in the following section.

5. Image schema combinations with the *scale* factor used in linguistic analysis: some examples

Scalarity in image-schema interaction is best seen in individual lexical items that denote actions, qualities, or states varying in intensity. A prototypical case for this is FORCE and *scale* combinations, where the scale calibrates the degree of action such as hitting. In addition, the valence may be both negative or positive, with a 'median value' (in our case: example 1.d.) in which scalarity is switched *off* as it has no implication for the affective meaning of the lexeme, while in the other instances it is *on* bearing diversified significance to the overall conceptualization. The symbol <s> denotes a linguistic construction with image-schema intensification and <ms> a metaphorical such conceptualization. Please note that lexical meanings are always context dependent. Thus, while we consider the lexeme “hit” schematically neutral here, it would not need to remain so in differently compiled lexical sets:

1. a. *smash someone* <s><F+++>

   b. *thwack someone* <s><F++>

   c. *punch someone* <s><F+>

   d. *hit someone* <s><F0>

   e. *smack someone* <s><F->

   f. *cuff someone* <s><F-->

   g. *tap someone* <s><F--->

Scales may equally influence PATHS, yet, in language, assuming only a positive valence. The interaction of PATH and *scale* can be envisaged in the sense of the extent of the path traversed, where scalarity affects the conceptualization in the sense of perspective and ‘progression on the path’. Thus, the instances where PATH is marked by *scale* + indicate the initial segment of the line; PATH++ would suggest midway positions on the trajectory and PATH+++ the final stages. Supposedly, any negative values of *scale* would entail regression, but this is not conceivable without previous progression. Therefore, PATH- is not possible. Rather, when necessary for the interpretation, we specify the directionality of the movement with annotations such as PATH+, spec: *up* or PATH++, spec: *reverse* (this is sometimes relevant in music analysis too). The situation is comparable with the schema BALANCE, where, in language, there can be no discretization of positive values (this is possible in music, though). It is either absolute or there is no balance in the true sense, which implies that BALANCE can be only affected by negative valence, indicating smaller of greater degree of balance disruption.

In our understanding, it is only the complex of two image schemas, namely FORCE and PATH, accompanied by the pronounced *scale* parameter that could refer to speed, since it is only with the increase of FORCE that the PATH can be "covered" with greater intensity, i.e., the speed of motion through space is greater. Rapidly passing over a longer stretch of path without greater force engaged in the process is difficult, if not impossible, to envisage. Just like musical tempo, speed, however, remains outside of the scope of this article.

Thus the onset point of a path would be equivalent to 0, whereas each major step of the way in terms of progression would amount to the value of <+>, and the final stage of reaching the end of the path would equal <+++>.

2. a. *The match is about to finish, still no goals.* <ms><P+++>

   b. *The third quarter of the match is on.* <ms><P++>

   c. *The match is well underway.* <ms><P+>

   d. *The match is just about to start.* <ms><P0>

The examples under (2) present various stages of development of a sports event. In such a context, they are all construed as metaphorical (<ms>) as the match is not physically moving anywhere, and the path-based progression is meant to stand for the time elapsed since the beginning of the event.

The instances in (3) present further complexity in the system, as they all exhibit a co-engagement of the PATH and FORCE schemas in interaction with CONTAINMENT with different implications of the *scale* factor for the linguistic expression, where the various positions of the conceptualizer can be revealed.

3. a. *troops go into Sievierodonetsk* <s><P+><F+><C+>

   b. *troops invade Sievierodonetsk* <s><P+><F++><C++>

c. *troops storm Sievierodonetsk* <ms><P++><F+++><C++>

The first two cases are arguably literal, and the third metaphorical, as no physical storm is involved in the event. Yet as the intensity of PATH, FORCE and CONTAINMENT progresses from the first to the third situation, both the scope of conceptualized occupation and the affective load of the semantic interpretation increase.

LINK, too, may be scalar, denoting the amount of physical or metaphorical connection (4):

4. a. John and Mary are *worlds apart*. <ms><L--->

b. John and Mary are *at opposite poles*. <ms><L-->

c. John and Mary are *separated.* <s><L->

d. John and Mary are *together*. <s><L0>

e. John and Mary are *attached*. <ms><L+>

f. John and Mary are *inseparable*. <s><L++>

g. John and Mary are *glued to one another*. <ms><L+++>

Once again, the conceptualizations with strong scalarity, either positive or negative, tend to be more metaphorical and emotionally laden (“worlds apart” and “glued to one another”).

A similar situation may apply with CONTAINMENT, though the type of scalarity here is somewhat different. Namely, this schema presupposes “a trajector through space to a particular endpoint configuration” (Mandler 2005: 150). Moreover, as Dewel (2005: 373−374) suggested: “it seems unlikely that a child’s earliest image schemas related to containment will be pure static relations in timeless space [...] It is much more likely that the earliest image schemas will involve activities and paths.” Thus, unlike combinations with FORCE, the *scale* here does not suggest intensity, but rather the *relation* of the Trajector to the container (5):

5. a. Mike is *out of his mind.* . <ms><C-->

b. Mike is *going out of his mind.* <ms><C->

c. Mike should *state his mind.* <ms><C0>

d. Mike is *returning to his right mind.* <ms><C+>

e. Mike is *in his right mind.* <ms><C++>

Obviously, the five statements are metaphorical, with the pronounced negative or positive containment in the first and the last likely profiling strong emotional involvement.

The remainder of the section presents several interesting cases from the set of 100 news headlines we have gathered from the online portals of the *New York Times* and the *Guardian* between 27 May and 1 June 2022. Newspaper headlines make a specific type of writing which bears distinctive qualities. They have more readers than any other type of text (Mardh 1980:11) and their "general impact is stronger than that affecting people who buy the newspapers, since they are often glimpsed almost everywhere" (Marcoci 2014). In other words, many more people are exposed to headlines than to full texts following those headlines. Such short texts provide condensed messages packed with meaning, stylistic creativity and communicative intent. Inadvertently or not, they reveal the major viewpoints of the author and establish the general tone of the article. As they can anticipate the content of the text to follow, they may also hint at the overall disposition of the writer toward the topic or the main focus. Moreover, they are often meant to appeal to the reader and grasp their attention, while simultaneously providing a particular framework of interpretation. Since it is imperative for the authors to 'catch the reader's eye' with considerable brevity in expression, this may be conducive to the extensive use of conceptual and orientation metaphors, a core interest of this paper. Being thus structured, headlines can be perceived as good indicators of the affect engaged on the part of the author manifest in the selection of the lexis, sentence construction (Fairclough 1995:21), presuppositions made, rhetorical tropes and the narrative (cf. Richardson 2007: 46-74).

On the present occasion, we concentrate on motion verbs and expressions extracted randomly from the corpus:

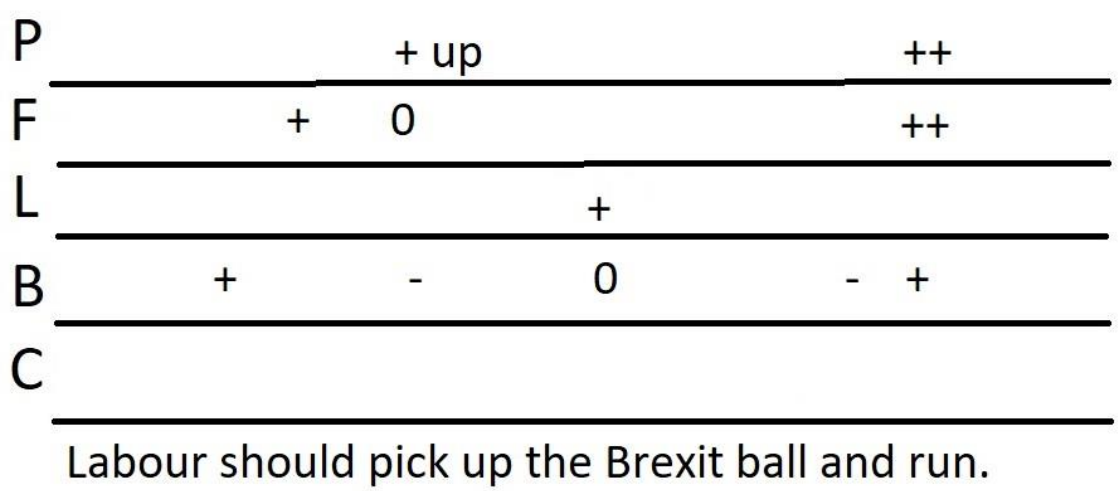
6. *Labour should pick up the Brexit ball and run*

<ms><FORCE0><PATH+[up]><ms><FORCE++><PATH++>

The sentence presents an unmarked schematic FORCE involved in picking up the object, lightly pronounced PATH, specified as going upward, and the act of running which requires a more substantial FORCE and implies a lengthier PATH. The entire expression contains three metaphorical expressions forming a cluster, of course, as the act of picking up the ball and running maps onto the proposed course of political action for the Labor party. The analysis with separate lines for the five schemas of interest provides a more detailed insight (Figure 8).

Figure 8.

A Dynamic Schema Analysis of the "Brexit" Sentence.



On this presentation, CONTAINMENT is not profiled (though it can be assumed from context, where the British political arena is metaphorized as a rugby court), LINK involves just the point of contact between the actor (metonymical of the Labor party) and the ball, while FORCE and PATH are present in both instances of movement: picking up the ball and, more strongly, running (away with it). The case of BALANCE is interesting: though very implicitly, the actor may be assumed to lose a bit of balance in the process of picking up the ball, and restore it during the process of running. In all, the “run” lexeme provides for the strongest schematic score: three schemas present, all with a positive valence, and the total numerical value of +5. Yet the major reason for the centrality of the verb “run” in the expression is that it establishes a *cumulative* schematic effect. For one, this verb is the principal factor of causation (in the sense of initiating an event; this is itself a strongly metaphorical notion according to Lakoff and Johnson 1999: 224). Yet more importantly, the image invoked by calling the verb is partly composed from the result of the employment of the previous schemas: one first needs to locate an [OBJECT], pick it up [FORCE, PATH, spec: *up*], and then run [FORCE++ PATH++]. The running is at the very least more durable than the act of picking something up, and it additionally progresses with this previously collected, metaphorical object in one’s hands. The aggregate effect strengthens the affectivity and metaphoricity of the resultant mental simulation. The conceptualization of such running is, therefore, most likely the point of the strongest affective impact on the reader, and the moment at which he or she should be certain that the sentence is to be interpreted metaphorically. Additionally, the proposed affective impact of schema cooccurrence and their metaphorical extensions is in line with previous work in the field dealing with the affective impact of metaphors in discourse (e.g., Cameron 2007a; 2007b; Cameron & Stelma 2004).

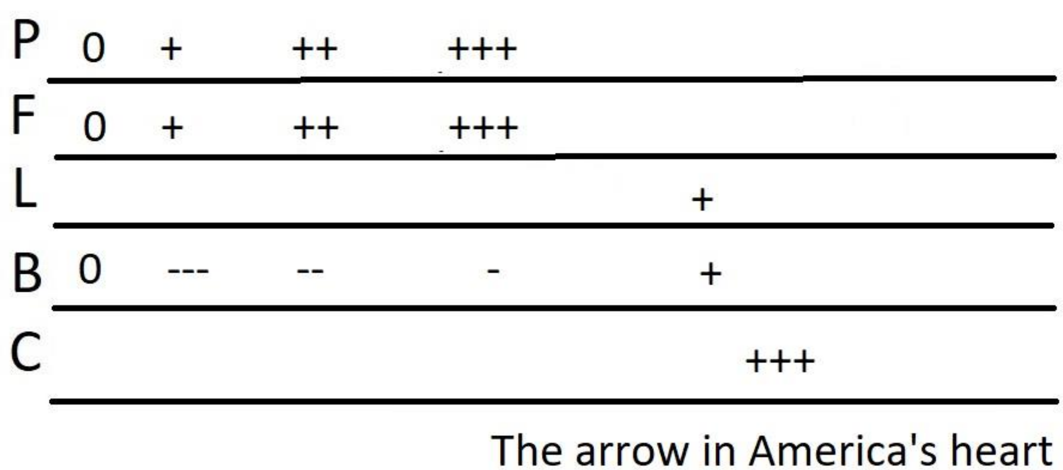
7. *The arrow in America’s heart*

   <ms><LINK+><CONTAINMENT+++>

This example is very interesting as its implied schemas are equally important for the interpretation as the visible ones. There is two-level metaphorization involved: the agent's body standing for America, and the heart working as a metonymy for a most precious set of values in this body/country. The arrow and the heart are linked physically, since they are touching one another, and the LINK is specified in such a way that the arrow is fully contained in the heart. The five-schema graph reveals more subtlety behind the expression (Figure 9).

Figure 9.

A Schema Analysis of the "Arrow" Sentence.



The strongest schematization, therefore, *precedes* the statement given in the headline. Hearing of the arrow in the heart, the reader will most likely assume that this arrow was first shot, which involved a combination of increasing PATH and FORCE and strongly decreasing (after the ejection) and then increasing BALANCE. The peak affective moment in the expression is the understanding that the arrow is now in the heart. This is first understood as the *result* of the entire operation. In addition, three simultaneous schemas, with positive valence, and the sum of

+5 most likely make this the strongest affective point in the expression, and the one with the highest metaphorical potential.

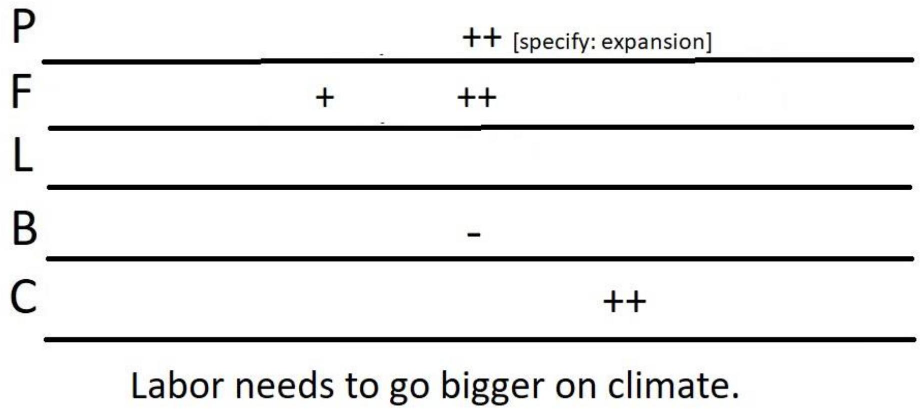
8. *Labour needs to go bigger on climate*

<ms><FORCE+><PATH++><CONTAINMENT+>

This metaphorical headline presents the Labor party as an agent that needs to "go bigger". The required physical expansion is presumably meant to metaphorically map onto an urge for the party to become much more engaged with issues of climate change. "On climate" is again metaphorical, as it presumes a stand *with regard to* the problem of climate, but is represented here as a specific type of CONTAINMENT, in which the "climate issue" sets the physical background – or surface –for the process of expansion. "Go bigger" is equally interesting, as it involves a PATH that does not progress linearly, but rather diffusely, in all directions, where the agent ultimately changes their shape through the process of moving. The comparative form of the adjective arguably justifies the increased scalarity here. The shape change of this kind may be an indicator of an even broader, higher-order principle, in which the PATH schema, combined with *scale*, can assume various kinds of directionality: upward, leftward, in a circle, expansion, thickening, and so on (for a similar phenomenon with pitch conceptualization in music, see Antović, Mitić & Benecasa 2020). A graphic presentation follows in Figure 10.

Figure 10.

A Schema Analysis of the "Climate" Sentence.



As seen in the graph, force activates already after the modal verb “need” and increases through the process of growing bigger. Here, of course, the PATH specified as expansion increases, as well. The BALANCE is arguably slightly lost at the beginning of this growth, after which it is restored. Thus, the conceptualization of “going bigger” against the background of “climate” provides for the strongest schematic intensity and affective / metaphorical interpretation.

9. *Progressives take aim at Democratic leadership over support for centrist candidates*

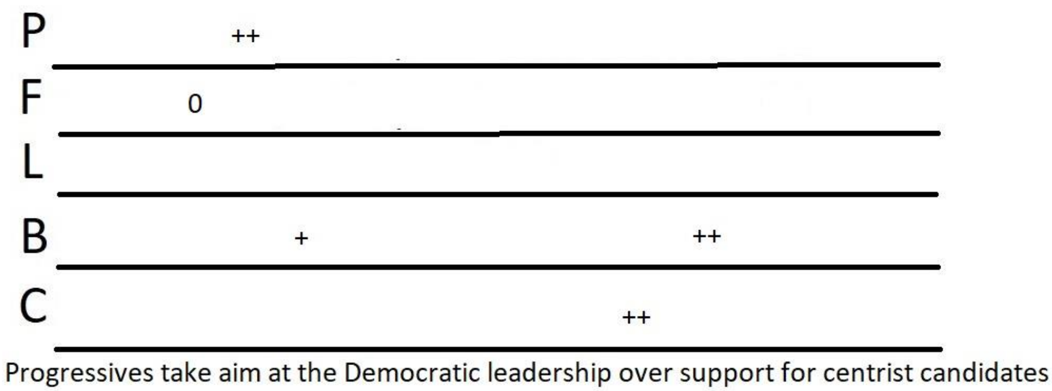
<ms><FORCE0><PATH+><CONTAINMENT++><BALANCE++>

The final example in the present paper is again fully metaphorical – ‘aiming’ at the leadership means that the progressives intend a political attack on the heads of their party because they appear to support more traditionalist candidates than the progressives would like. The aiming involves a (rather neutral) FORCE to produce the movement through a PATH, the prepositional phrase starting with ‘over’ works as a background, providing CONTAINMENT for the entire statement, while the ‘support’ phrase ensures a strong metaphorical balance for the selection of the centrist candidates. (The ‘centrists’ themselves are a manifestation of the

CENTER/PERIPHERY schema, which is not of direct interest in the present paper). A possible graphic representation would look as in Figure 11.

Figure 11.

A Schema Analysis of the "Aim Taking" Sentence.



Again, it appears that the strongest emotional effect and metaphorical interpretation relates to the moment of "taking aim at the leadership", as it involves the interaction of FORCE, MOVEMENT and a restored BALANCE upon the end of this movement (apparently, while just "taking aim", the acting agent produces an imbalanced movement in all directions; this stabilizes upon locking the target). Accordingly, the interaction of these schemas suggests an affective and associative interpretation in addition to the mere understanding of the denotation of the headline.

6. Conclusion

The present text is to be understood as a programmatic paper of a conception in the fundament of a research project that incorporates language, music and visual cognition. We hope to have made a statement about the importance of underlying image schema complexes in conceptualizations relevant to language and music. Moreover, we have provided a formal ground for the instances where image schema interactions based on the higher-order factor *scale* can be found to produce a range of specific conceptualizations varying in semantic contingency and pragmatic effect. As stated, musical and linguistic analyses result in specific types of advantages and drawbacks. In music, schemas are relatively easy to detect formally from the very elements employed in the notation; on the other hand, due to all-present recursion, schema valences and intensities can differ substantially depending on the size of the musical excerpt one is focusing on (e.g., a segment, musical sentence, or entire movement). This is why, for the time being, we propose a focus on smaller musical segments (up to eight measures) and local schematic effects (e.g., FORCE inferred from the musical dynamic alone). For the same reason, we limit our interest to the common-practice period of homophonic Western art music and leave broader cross-cultural comparisons for future work. In language, the detection of schemas depends on interpretive processes, and is somewhat arbitrary; equally, schematic dynamicity here is meant to capture *events* described by the linguistic material, not the language structure itself. Yet, once this stage is complete, the analysis of schematic interactions quite naturally suggests which parts of the structure are likely to produce the strongest affective responses by the listeners and/or be interpreted as metaphorical. To avoid arbitrariness as much as possible, in this process we propose that researchers work together in identifying the schematic complexes through a well-defined algorithm: defining criteria for classification, annotating corpora in pairs of researchers, checking issues of disagreement with a third researcher, organizing a final majority vote in cases that remain ambiguous even after this stage.

We hope that the amended approach to schematicity that we are proposing here introduces the following novelties: it studies (1) image schemas in *interaction* (specific complexes of at least three schemas in both music and language), with (2) individual schemas defined by means of *formalized intensities* and, where appropriate, *valences*, (3) a corresponding discretized system which would be able to *predict* (and also, experimentally test) to what extent such variation of parameters changes the affective valence and arousal in participants' cross-modal, referential descriptions, and (4) possible commonalities in such a response to both linguistic and musical stimuli on the basis of the same image-schema complexes and intensities, and irrespective of the modality in which they are manifested.

Much more work is needed for substantial progress, of course. In its current form, our proposal is only a tentative method of analysis that awaits further testing with more, and more varied, language data. Its hopeful contribution is also its most obvious limitation. Given that musical and linguistic semantics show strong ontological differences – language meaning being mostly interpretive, and musical one iconic – the relatively intuitive, perhaps mechanistic method for extracting schemas from musical structure becomes much more difficult in language. There, prior to ascribing schemas, one already needs to have used the linguistic utterances to *interpret* how entire scenes are simulated in the mind of the interlocutors. In this latter process, the employment of an intermediary – a human informant, a dictionary, an AI parsing routine, or perhaps all of these together – cannot be avoided. That in turn makes the analysis more difficult to control and liable to various interpretations. But one can hardly avoid such a methodological problem in most research on affective meaning in language.

From the musical angle as well, splitting the stimulus into different parameters (melody, harmony, rhythm, dynamics…) for settling observable measures within a complex musical situation poses problems of its own. Extracting schemas from a musical *score* is one thing, and extracting them from the ongoing aural experience quite another. This is an old point of disagreement among music cognition practitioners, and one that needs to be addressed in future work.

These issues irrespective, as we are progressing with our work on musical and linguistic corpora, we hope this first program paper has suggested a way in which the powerful concept of the image schema from cognitive linguistics can be further developed so as to be applicable in formal and computational contexts.

Acknowledgement

This research was supported by the Science Fund of the Republic of Serbia, Grant No. 7715934, Structuring Concept Generation with the Help of Metaphor, Analogy and Schematicity – SCHEMAS.

References

Antović, Mihailo. 2009. Musical metaphors in Serbian and Romani children: An empirical study. *Metaphor and Symbol* 24(3). 184–202.

Antović, Mihailo. 2018. Schemas, grounds, meaning: On the emergence of musical concepts through conceptual blending. *Musicae Scientiae* 22(1). 57-71.

Antović, Mihailo. 2022. *Multilevel grounding: A theory of musical meaning*. Abingdon and New York: Routledge.

Antović, Mihailo, Jana Mitić & Naomi Benecasa. 2020. Conceptual rather than perceptual: Cross-modal binding of pitch sequencing is based on an underlying schematic structure. *Psychology of Music* 48(1). 84–104.

Athanasopoulos, George & Mihailo Antović. 2018. Conceptual integration of sound and image: A model of perceptual modalities. *Musicae Scientiae* 22(1). 72-87.

Barrett, Lisa Feldman. 2006. Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review* 10(1). 20–46.

Barsalou, L. W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22(4). 577-660.

Bernstein, Leonard. 1976. *The unanswered question: Six talks at Harvard*. Cambridge, MA: Harvard University Press.

Bierwisch, Manfred. 2009. Sprache—Musik—Bild. *Zeitschrift für Literaturwissenschaft und Linguistik* 39(3). 8–34.

Brower, Candace. 2000. A cognitive theory of musical meaning. *Journal of Music Theory* 44(2). 323–379.

Cameron, Lynne. 2007a. The affective discourse dynamics of metaphor clustering. *Ilha do Destero Florianopolis* 53. 41–61.

Cameron, Lynne. 2007b. Patterns of metaphor use in reconciliation talk. *Discourse and Society* 18(2). 197–222.

Cameron, Lynne J. and Juurd H. Stelma. 2004. Metaphor clusters in discourse. *Journal of Applied Linguistics* 1(2). 107–136.

Cienki, Alan. 1997. Some properties and groupings of image schemas. In Marjolijn Verspoor, Ewe Sweetser & Kee Dong Lee (eds.), *Lexical and Syntactical Constructions and the Construction of Meaning*, 3–15. Amsterdam: John Benjamins.

Clausner, Timothy & William Croft. 1999. Domains and image schemas. *Cognitive Linguistics* 10(1). 1–31.

Dewel, Robert. 2005. Dynamic patterns of CONTAINMENT. In Beate Hampe (ed.), *From Perception to Meaning. Image Schemas in Cognitive Linguistics*, 369–394. Berlin and New York: Mouton de Gruyter.

Dolscheid, Sarah, Shakila Shayan, Asifa Majid & Daniel Casasanto. 2013. The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science* 24(5). 613-621.

Duncan, Seth & Lisa Feldman Barrett. 2007. Affect is a form of cognition: A neurobiological analysis. *Cognition and Emotion* 21(6). 1184–1211.

Eitan, Zohar. 2017. Musical connections: Cross-modal correspondences. In: Richard Ashley and Renee Timmers (eds). *The Routledge Companion to Music Cognition*, 213-224. New York: Routledge.

Eitan, Zohar & Renee Timmers. 2010. Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition* 114(3). 405-422.

Evans, Vyvian & Melanie Green. 2006. *Cognitive Linguistics: An introduction*. Edinburgh: Edinburgh University Press.

Fairclough, Norman. 1995. *Media Discourse*. London: Arnold.

Grady, Joseph E. 2005. Image schemas and perception: Refining a definition. In Beate Hampe (ed). *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, 35–55. Berlin: Walter de Gruyter.

Gray, Elizabeth, K. & David Watson. 2007. Assessing positive and negative affect via self-report. In James A. Coan & John JB Allen (eds.), *Handbook of Emotion Elicitation and Assessment*, 171–183. Oxford: Oxford University Press.

Hampe, Beate. (ed.) 2005. *From perception to meaning: Image schemas in cognitive linguistics*. Berlin: Walter de Gruyter.

Hedblom, Maria, Oliver Kutz & Fabian Neuhaus. 2015. Choosing the right path: Image schema theory as a foundation for concept invention. *Journal of Artificial General Intelligence* 6(1). 21–54.

Hedblom, Maria, Oliver Kutz & Fabian Neuhaus. 2016. Image schemas in computational conceptual blending. *Cognitive Systems Research* 39. 42-57.

Hedblom, Maria, Oliver Kutz, Rafael Peñaloza & Giancarlo Guizzardi. 2019. Image schema combinations and complex events. *Künstliche Intelligenz* 33. 279–291.

Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 2002. *Foundations of language*. New York: Oxford University Press.

Jackendoff, Ray & Fred Lerdahl. 2006. The capacity for music: What is it, and what's special about it?. *Cognition* 100(1). 33–72.

Johnson, Mark. 1987. *The body in the mind. The bodily basis of meaning, imagination, and reasoning*. Chicago: The University of Chicago Press.

Johnson, Mark. 2005. The philosophical significance of image schemas. In Beate Hampe (ed.), *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, 15–33. Berlin: Walter de Gruyter.

Johnson, Mark. 2007. *The meaning of the body. Aesthetics of human understanding*. Chicago: The University of Chicago Press.

Johnson, Mark & Steve Larson. 2003. "Something in the way she moves" – Metaphors of musical motion. *Metaphor and Symbol* 18(2). 63–84.

Koelsch, Stefan. 2012. *Brain and music*. Chichester: John Wiley & Sons.

Koffka, Kurt. 1936. *Principles of Gestalt psychology*. London: Kegan Paul.

Krzeszowski, Tomasz P. 1993. The axiological parameter in preconceptual image schemata. In Richard Geiger & Brygida Rudzka-Ostyn (eds.), *Conceptualizations and Mental Processing in Language*, 307–330. Berlin & New York: Mouton de Gruyter.

Küssner, Mats B., Liila Taruffi & Georgia A. Floridou (eds). 2022. *Music and Mental Imagery*. New York: Routledge.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: The University of Chicago Press.

Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: The University of Chicago Press.

Lakoff, George & Mark Johnson. 1999. *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.

Lakoff, George & Mark Turner. 1989. *More than cool reason: A field guide to poetic metaphor*. Chicago: The University of Chicago Press.

Lakoff, George. 1990. The invariance hypothesis: Is abstract reason based on image schemas? *Cognitive Linguistics* 1(1). 39–74.

Larson, Steve. 2012. *Musical forces.* Bloomington: Indiana University Press.

Lerdahl, Fred. 2001. *Tonal pitch space*. New York: Oxford University Press.

Lerdahl, Fred & Ray Jackendoff. 1983. *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Mandler, Jean & Cristóbal Pagán Cánovas. 2014. On defining image schemas. *Language and Cognition* 6(4). 1–23.

Mandler, Jean. 1992. How to build a baby: II. Conceptual primitives. *Psychological Review* 99(4). 587–604.

Mandler, Jean. 2004. *The foundations of mind: Origins of conceptual thought*. New York: Oxford University Press.

Mandler, Jean. 2005. How to build a baby: III. Image schemas and the transition to verbal thought. In Beate Hampe (ed.), *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, 137–163. Berlin: Walter de Gruyter.

Mandler, Jean. 2012. On the spatial foundations of the conceptual system and its enrichment. *Cognitive Science* 36(3). 421–451.

Marcoci, Sanda. 2014. Some typical linguistic features of English newspaper headlines. *Linguistic and Philosophical Investigations* 13. 708–714.

Mårdh, Ingrid. 1980. *Headlinese: On the grammar of English front page headlines*. Lund: CVVK Gleerup.

Oakley, Todd. 2007. Image schema. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics*, 214–235. Oxford: Oxford University Press.

Popova, Yanna. 2005. Image schemas and verbal synaesthesia. In Beate Hampe (ed.), *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, 395–419. Berlin: Walter de Gruyter.

Powers, Harold S. 1980. Language models and musical analysis. *Ethnomusicology* 24(1). 1–60.

Pratt, Carroll, C. 1930. The spatial character of high and low tones. *Journal of Experimental Psychology* 13(3). 278–285.

Richardson, John E. 2007. *Analysing newspapers: An approach from critical discourse analysis*. Houndmills: Palgrave MacMillan.

Rothfarb, Lee. 2002. Energetics. In Thomas Christensen (ed.), *Cambridge History of Western Music Theory*, ed. 927-955, Cambridge: Cambridge Univeristy Press.

Russell, James A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39(6). 1161–1178.

Russell, James A., Anna Weiss & Gerald A. Mendelsohn. 1989. Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology* 57(3). 493–502.

Russell, James A. 2003. Core affect and the psychological construction of emotion. *Psychological Review* 110(1). 145–172.

Ruwet, Nicolas & Mark Everist. 1987. Methods of analysis in musicology. *Music Analysis* 6(1/2). 3–36.

Saslaw, Janna. 1996. Forces, containers, and paths: The role of body-derived image schemas in the conceptualization of music. *Journal of Music Theory* 40(2). 217–243.

Spence, Charles. 2020. Assessing the role of emotional mediation in explaining crossmodal correspondences involving musical stimuli. *Multisensory Research* 33(1). 1–29.

Stadler, Michael W. 2020. *The ontological nature of part-whole oscillations*. Vienna: Austrian Academy of Sciences Press.

Talmy, Leonard. 1988. Force dynamics in language and cognition. *Cognitive Science* 12(1). 49-100.

Watson, David, Lee Anna Clark & Auke Tellegen. 1988. Development and validation of brief measure of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* 54(6). 1063–1070.

Zbikowski, Lawrence M. 2002. *Conceptualizing music: Cognitive structure, theory, and analysis*. New York: Oxford University Press.

Zbikowski, Lawrence M. 2008. Metaphor and music. In Raymond W. Gibbs (ed). *The Cambridge Handbook of Metaphor and Thought*, 502–524. Cambridge: Cambridge University Press. Zbikowski, Lawrence M. 2017. *Foundations of musical grammar*. New York: Oxford University Press.