The Serbian Association for the Study of English
THE SASE JOURNAL

https://doi.org/10.46630/sase.2.2026

University of Niš
Faculty of Philosophy

UDC 821.111 ISSN 3042-2930

THE SERBIAN ASSOCIATION FOR THE STUDY OF ENGLISH

THE SASE JOURNAL

Vol. 2, 2026

Editors
Jelena Danilović Jeremić
Marta Veličković

Niš, 2026

CONTENTS

PROJECT RESULTS IN TERMS OF IMPLICATIONS FOR TEACHING, EDUCATION OF SPECIAL GROUPS, AND RHETORICAL EXPRESSION

Science Fund of the Republic of Serbia
Program: IDEAS
Project *Structuring Concept Generation with the Help of Metaphor, Analogy, and Schematicity* (Project No. 7715934)
Acronym: *SCHEMAS*
Prepared by: Vladimir Ž. Jovanović, PhD, full professor, Vladan Pavlović, PhD, full professor, and Aleksandra Janić Mitić, PhD, associate professor

1. Introduction

The project bearing the acronym SCHEMAS deals with the ways in which image schemas – along with certain other parameters – as the building blocks of concepts in our minds can successfully be combined to produce more complex ideas in the mind. Similar principles of combination can be found in the "language" of metaphors, the "language" of music, or the "language" of visual sequences, among others. In the research underlying this project, we examined the process of concept formation in our minds – something that forms the basis of human cognition and thought, one of the fundamental capacities that distinguishes humans from other beings. We sought to uncover the dynamic mechanism that governs the creation of such "units" of thought when we speak figuratively, when we listen to or create music, or when we watch a sequence of images. By doing so, we can identify common elements across several different domains where thought processes manifest themselves – insights that could later be applied in efforts by experts to teach machines to communicate according to these principles, so that they can "understand" us just as we understand them.

The project is based primarily on the postulates of cognitive science and the theoretical framework of cognitive linguistics, although it also applies scientific tools from other related fields. During the study of the phenomena that were in focus, methodological procedures from corpus linguistics were used to examine large groups of metaphorical expressions, as well as musical phrases or visual sequences. In addition, members of the project team used psycholinguistic experiments in an

attempt to delve deeper into the human mind in order to uncover correlations and abstract principles that govern the creation of meaning in general − both in language and elsewhere.

The main goal of the research was to test and confirm the idea that image schemas, a term used to describe preconceptual constructs underlying our cognition, are not necessarily static in nature but can gradually develop, evolve, and interact with other schemas, thereby forming more complex conceptual structures.

The expected outcome of the work carried out within the SCHEMAS project on the aforementioned issues consists of clearly presented evidence − in several forms (through corpus analysis, theoretical considerations, experimental investigations, and more) − of the existence of mechanisms for the dynamic interaction of conceptual schemas (hence the project acronym). The project results have been presented to the wider academic community in the form of at least fifteen scientific papers published in reputable international journals and more than twenty presentations at international conferences.

From a scientific standpoint, the value of the project results can primarily be seen in those areas of cognitive science concerned with uncovering the patterns and characteristics of the human thought processes. The results show that the mechanism underlying such a dynamic relationship is not characteristic of language alone but represents a more general cognitive principle. At the same time, cognitive linguistics has been advanced to some degree by the recognition and confirmation that image schemas need not be viewed as static constructs, but rather possess developmental characteristics. Research in the field of multimodality has progressed thanks to results showing that the mechanism in question operates across three different systems (figurative language, music, and visual expression).

The concrete application of the project results lies in the possibility that everything produced as a contribution to the study of the given problem can, in various ways, assist creators of educational and pedagogical materials, as well as those working in other domains of science and creative practice. In the field of language education, the project results enhance our understanding of how concepts are acquired in the process of learning and teaching foreign languages in general. Based on the knowledge gathered through research and experiments, certain improvements could be expected in musical and audiovisual aids used in musical education and visual arts, especially in cases where special education is required. Furthermore, the field of rhetorical expression in communication could make use of some of our research findings to improve techniques for persuading others, particularly in the realm of propaganda messages in the media, politics, or economics.

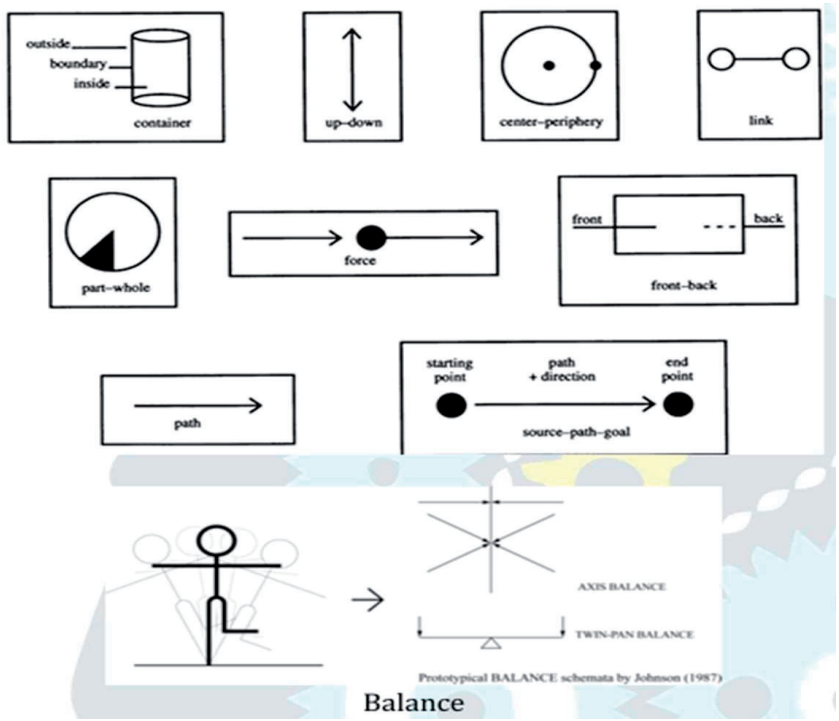
2. Image schemas

In cognitive linguistics, image schemas are viewed as specific subconscious, abstract patterns in the mind − conceptual sub-structures primarily based on our experience of spatial relations such as *inside–outside*, *up–down*, *near–far*, *center–*

periphery, *part–whole*, *front–back*; on the manipulation of physical objects (which involves force); and on movement through space (*source–path–goal*). In other words, the source of image schemas can be understood as our immediate sensorimotor and perceptual experience that arises from interaction with the external world.

The importance of image-schematic patterns lies in the idea that they underlie the concepts we use and therefore form the basis of human thought − an idea that goes back at least to the work of Immanuel Kant, that is, at least to the 18th century, if not much earlier. By studying them, we can better understand human thought processes and communication.

Some typical representations of image schemas are shown in the diagram below (the diagrams and images mentioned in this brochure were taken from Antović, 2024; Johnson, 1987, as well as from freely available online materials; translations of the English terms used in those diagrams are provided in the text below).



3. Image schemas and metaphoricity

What is of exceptional importance here is the insight that image schemas can also be viewed as fundamental elements of metaphor. In other words, the physical relationship represented in a schema can serve as an experiential basis for understanding abstract concepts. For example, a physical object may be located

inside or outside the boundary of a container. A key may be in a drawer, students in a classroom, a ball in a cupboard, a sofa in a room, soup in a pot; someone may be leaning on the outer wall of a house; glasses may be on or beside a printer; a cat may be in the basement, etc. In all these cases, there is a container − fully or partially bounded by its boundaries (drawer, pot, house, printer, basement) – in relation to which other physical objects (key, students, ball, sofa, etc.) occupy some position.

However, many abstract concepts are structured in a similar way. Just as a key can be in a drawer, an event can occur *in March*, or *in August*, or *in December*. In other words, non-physical − abstract − concepts, in this case those relating to temporal units (months of the year), are unconsciously understood as *containers within which* individual events can be located. Similarly, someone may be *in a mentally difficult state*, or *in danger*, *on the threshold of a scientific discovery*, *under great (psychological) pressure*, *in a (romantic) relationship*, etc.

Likewise, the action of a force can be understood both literally and figuratively (e.g., *The flood weakened the foundation of the house* and *Such a decision undermined the unity among the members of that association*).

The same applies to the concept of balance − a force such as an earthquake may *disrupt the balance of a building*, just as an event may, figuratively speaking, *throw a person off balance*.

The same holds for the concept of a *link* or *connection*. For instance, a twig of an oak tree used as a *badnjak* may be tied to a bit of straw with a string, just as two people may be emotionally connected.

However, what particularly stimulates interest in image schemas within cognitive science is the fact that they not only form the basis of a large part of language, that is, not only appear in linguistic material, but can also be manifested − and serve as tools of conceptualization − in numerous other domains of human behavior and activity.

For instance, people, based on everyday experience, come to the conclusion that what is *up* is generally *good*, while what is *down* is generally *bad*. For example, a farmer who harvests a very large quantity − a heap − of wheat or other grains associates that large quantity with something *good* − he knows that with these grains he will be able to feed both himself and his domestic animals, such as poultry or pigs, and perhaps even sell part of the yield to earn income. Conversely, if the heap of grain obtained after the harvest is relatively small/low, this would threaten both his own survival and that of his domestic animals. In this case, what is *small* or *low* is experientially associated with something *bad*. Similarly, if a person is healthy and upright, that person can easily run and engage in various activities, such as farming and animal husbandry, and generally be active in different areas of life. Therefore, uprightness (i.e., the *up* position) is experientially associated with something good. Conversely, if a person is ill, even temporarily, that person is typically confined to bed, in a *down* position, which, based on experiential reasons (typically unconsciously), is associated with something bad (hence expressions like *fall into a coma*).

This association can be observed not only in language − when, for example, various social relationships are structured according to the pattern *up* = *good*, *down*

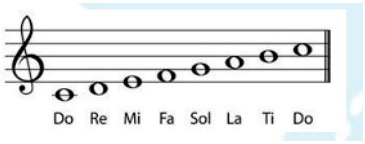
= *bad* (e.g., *He is at the head of the company*, *She has advanced high on the social ladder*, *He has fallen on hard times*) – but also in many other forms of human behavior that are not necessarily linked to language. For instance, a monarch's throne is typically placed on an elevation, even if only a few steps higher, rather than in a depression or a pit. Similarly, respect for the deceased is shown by looking downward rather than upward, sometimes by kneeling (e.g., at a grave). Executive offices in an organization are typically located on higher floors, often on the top floor or in a distinct elevated section. Depictions of Jesus Christ, and later of the apostles, are typically placed in the highest positions (at the top of the iconostasis, on the top of a dome, etc.) in Christian religious buildings, as illustrated in the images below (Antović, 2024; Rasulić, 2004).

Thus, the *up–down* schema, like all other schemas, is applied to numerous metaphorical or figurative concepts that share the same abstract structure. In other words, our cognitive system reduces the perception and proprioception of human bodily activities to a rudimentary image-schematic form of relation, which is then manifested not only in language but also across a wide spectrum of human action.

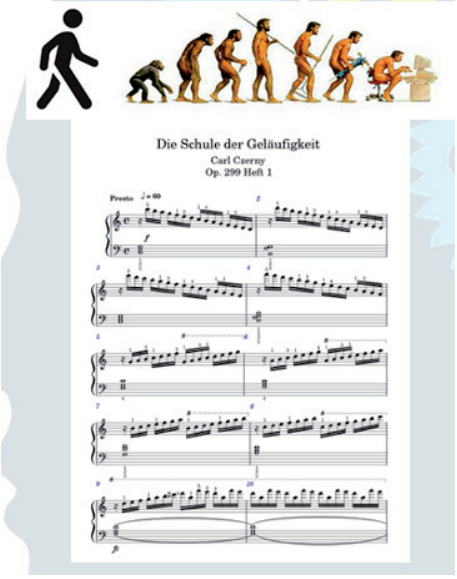
4. Image schemas in music and other fields of human thought and activity

Our goal in the following sections will be to focus in particular on how the insights mentioned above, as well as some additional similar insights that we will also present, can be applied, especially in the field of music. At the same time, we will continue to refer to examples that do not necessarily belong exclusively to music (or language), but also to various other forms of human activity, in an effort to emphasize once again the pervasive presence of image schemas in the patterns of human thought and action across different spheres of life.

For example, the aforementioned *up–down* schema is, of course, projected onto vertical systems of musical notation and perception, significantly contributing to the ubiquitous metaphor of musical relationships and motion.

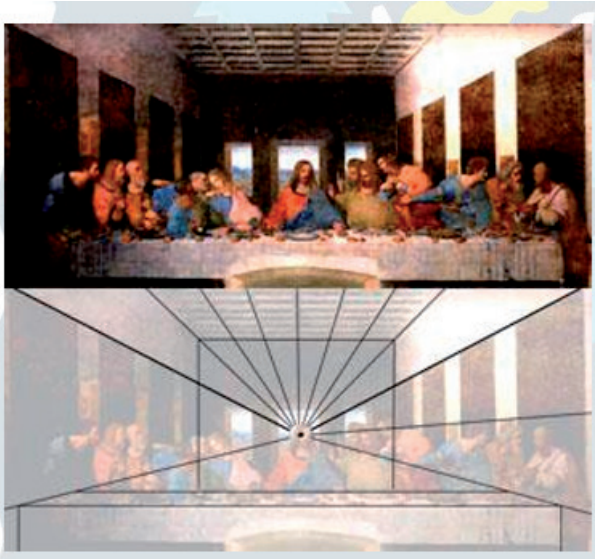


Similarly, at the core of the *path* schema lies the physical traversal of actual distances, which is then mapped onto numerous forms of human activity. For example, it can be applied to how we perceive the path of human evolution, or, in music, to how musical passages are organized on the staff, such as in the études of Carl Czerny (1791–1857), for instance his composition *Die Schule der Geläufigkeit* ("School of Velocity"), illustrations of which are provided below.



Finally, the importance of the *center–periphery* relationship may arise from the significance of the central parts of our body for survival, such as the heart, and, in that sense, the comparatively lesser importance of certain peripheral parts of the body (such as nails).

The concept of centrality versus peripherality of objects is also important metaphorically, as can be seen in the way the famous painting *The Last Supper* by Leonardo da Vinci is organized.



In music, of course, this is reflected in the hierarchy of pitch and the centrality of certain chords relative to others, for example, the tonic triad in relation to the dominant.

The project within which this brochure is being produced adds another aspect to these well-known concepts: the idea that schemas can be dynamically scaled in intensity and in real time (Antović, Jovanović, Figar, 2024; Pavlović, Janić Mitić, Mitić, 2024).

Here, we will briefly show how this can be applied to the schemas of *force*, *path*, *balance*, *connection*, and *containment*, first in language, before returning to music.

For instance, the relative intensity of force can be represented by the symbol F and a combination of minus and plus signs. Examples from language, arranged from least to greatest perceived force, are as follows: *touch/tap* <F--->, *pat* <F->, *splash* <F->, *hit* <F>, *strike strongly* <F+>, *strike with full force* <F++>, *reduce to dust and ashes* <F+++>.

For scaling the *path* schema, the length of the distance covered can be indicated (typically without negative values), for example: *The EU entry is uncertain* <P>, *Serbia has made some progress toward the EU* <P+>, *On the way to Brussels* <P++>, *The letter has arrived* <P+++>.

For scaling *balance*, one can indicate its mere presence, as well as a greater or lesser degree of disruption, if any, as in the following examples: *It is a delicate act of balancing* <B>, *This may lead to destabilization of the country* <B->, *Their relations were particularly disrupted after…* <B-->, *Polarization due to the war in Ukraine is enormous* <B--->.

For scaling *connection/linkage*, one can indicate the closeness or distance (physical or metaphorical) between two or more entities, as in: *Jovan and Marija are two different worlds* <L--->, *The country has ended military cooperation with Russia* <L-->, *If he were to hand over the illegally obtained money…* <L->, *They are together* <L>, *They became very close* <L+>, *They became inseparable* <L++>, *This connected them inseparably* <L+++>.

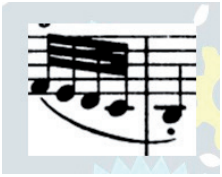
For scaling *containment*, one can indicate the degree to which an entity is located within a physical or metaphorical container, as in: *He took the documents out of the drawer* <C--->, *That country will be integrated into the EU* <C+++>.

In language, such intensities are interpretive; that is, drawing conclusions about the strength of an image schema requires referential semantic processing. In other words, from experience, we need to understand that, in the example *Jovan and Marija are two different worlds*, a world is an object that is inherently very large. Therefore, we interpret the example by recognizing that not only is there no connection between Jovan and Marija, but also that their distance − whether emotional or in some other sense − is extremely great.

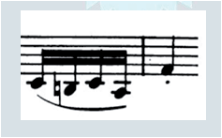
In music, we find a similar phenomenon, but it is easier to locate because it can be observed in the formal structure of the musical notation or in the parameters used to measure the characteristics of the sound stimulus.

Thus, in music, L- suggests a small change from a more connected to a less connected articulation in terms of performance technique, for example, when *legato*

(tones played smoothly without interruption) transitions to *portato* (tones that are slightly shortened and played in a somewhat detached manner). Such a transition is illustrated in the diagram below.



On the other hand, L-- indicates a much more noticeable change, for example, from *legato*, as described above, to *staccato* (where the actual duration of each note is significantly shortened and each note seems to represent an individual sound impulse). This, of course, can be easily observed in the score based on the notational symbols or measured from the sound stimulus. This transition is also illustrated in the diagram below.

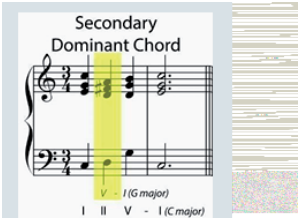


Similarly, the intensity of force − that is, the strength/loudness/intensity of the performance of a musical piece − can be determined. As a reminder, the categories of performance intensity and the corresponding standard dynamic markings are as follows:

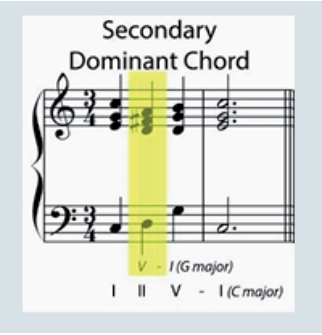
- as soft as possible (Italian: *pianissimo possible*, *ppp*)
- very soft (Italian: *pianissimo*, *pp*)
- soft (Italian: *piano*, *p*)
- moderately soft (Italian: *mezzopiano*, *mp*)
- moderately loud (Italian: *mezzoforte*, *mf*)
- loud (Italian: *forte*, *f*)
- very loud (Italian: *fortissimo*, *ff*)
- as loud as possible (Italian: *fortissimo possible*, *fff*)

In our notation, this gradient from the softest to the loudest possible performance intensity would range from <F---> to <F+++>.

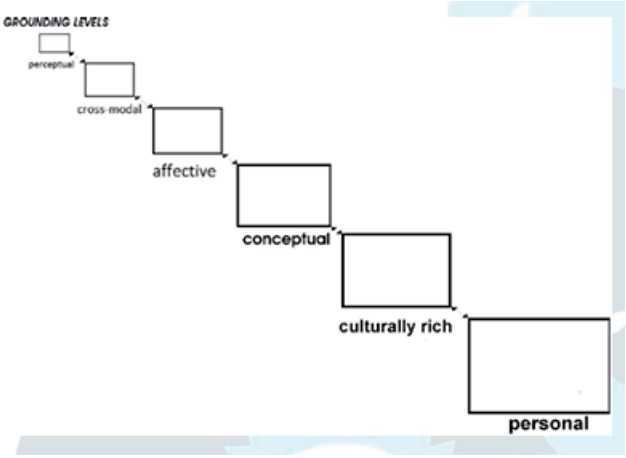
Additionally, in music, one can also speak of *path*, that is, a greater or lesser number of notes played per unit of time. For example, a leap of one octave, as in the example below, could be represented as <P++>.



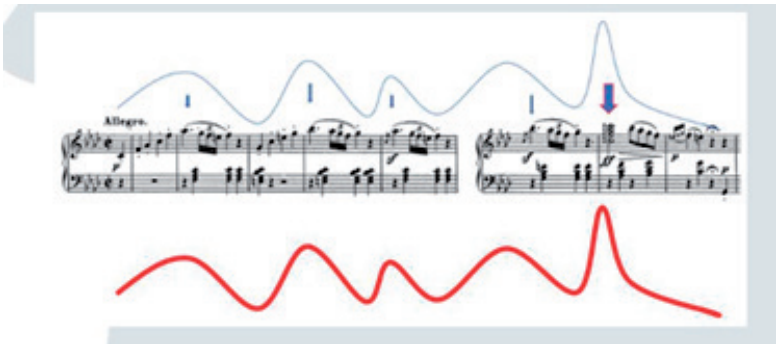
In music as well, one can also speak of *balance*, that is, its lesser or greater disruption. For example, in the illustration below, we can see one consonant chord and one strongly dissonant chord, which is the central (tonic) chord and the secondary dominant, which in our notation can be represented as <B->.



Broadly speaking, this schematic understanding of music constitutes only one layer of musical interpretation. Namely, Antović (2022) introduces six hierarchical and partially recursive (repeating) levels of grounding musical meaning: the formal-perceptual, the aforementioned image-schematic, the affective, the conceptual, the cultural, and the personal/individual levels, as illustrated in the diagram below.

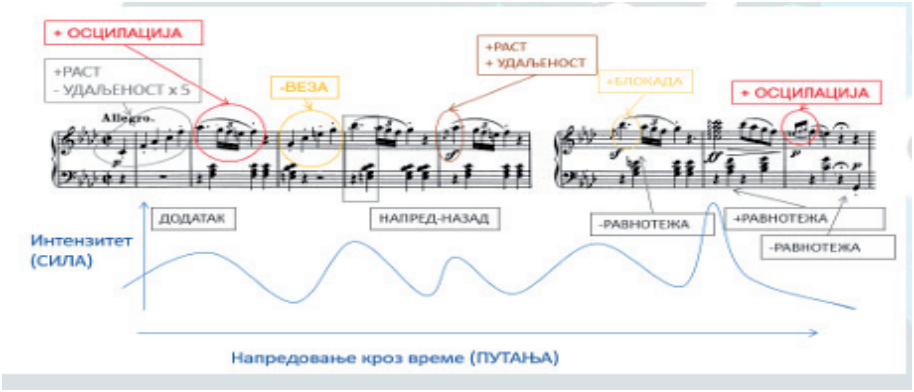


As an example, we can take the beginning of Beethoven's first piano sonata in F minor. We can then ask how the process of meaning construction unfolds in this passage. It certainly seems reasonable to first observe the levels of formal energy change derived from the sound stimulus. In other words, one can initially identify the relative peaks in the structure, including the final peak of the passage, which we will analyze below from the musical work, showing the strongest musical energy.



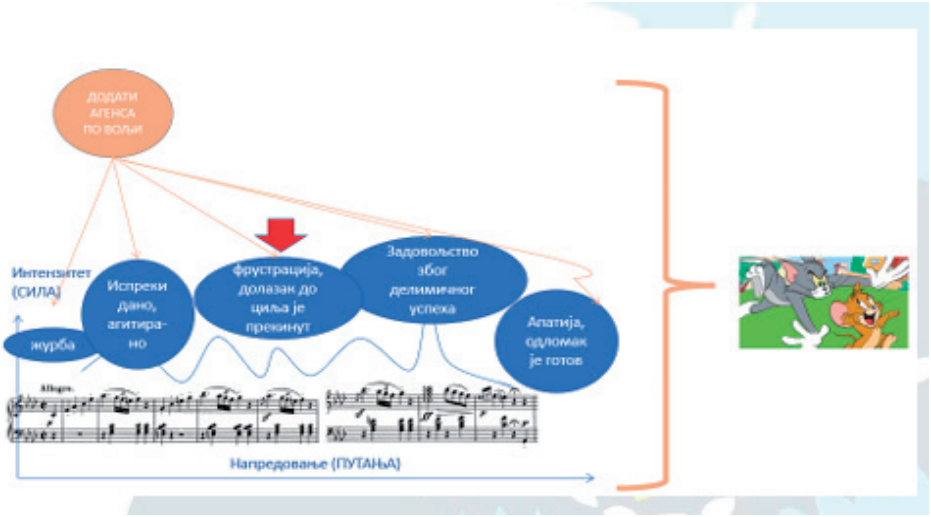
In the next step, which concerns interpretation based on compositionally immanent image schemas, a distinction can be made between higher- and lower-order phenomena. At the highest level, viewed horizontally, the entire musical passage progresses through time, that is, it has its flow, creating an image-schematic *path*. Viewed vertically, one can say that the musical flow rises and falls, that is, increases and decreases between points of greatest energy. This establishes the concept of musical *intensity* or *force*, which refers not only to simple forte articulation but also to more or less tense harmonic connections or articulatory intensities. This approach can be considered the first step toward making the lived musical experience easier to interpret.

Then, at a more specific level, in the given musical passage, one can identify specific image schemas such as *elevation*, *verticality/rise*, *distance*, *oscillation*, *present* or *absent connections*, *blockages*, *supports*, and *gestalts of forward and backward movement*, as illustrated in the diagram below for the same musical passage.

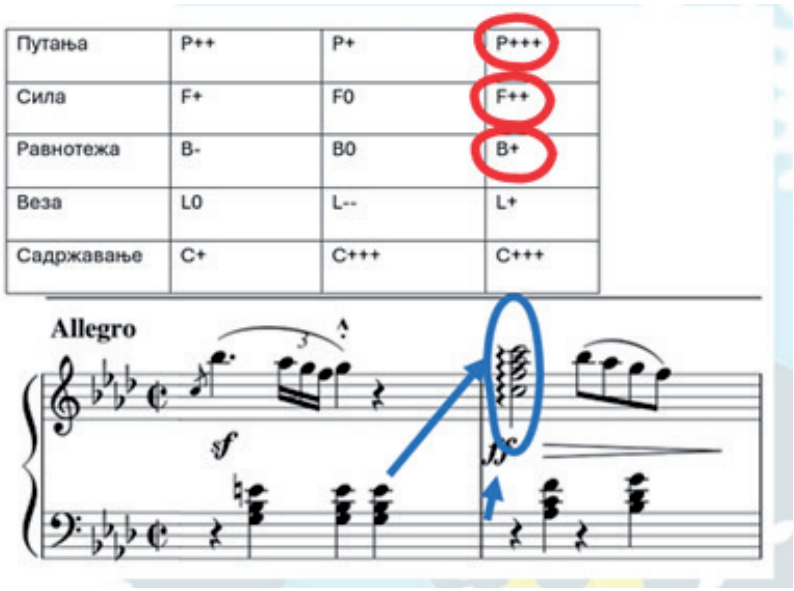


In the next steps, which concern the affective, conceptual, cultural, and individual levels, one can reflect on the affective and emotional, conceptual, and social aspects of meaning construction related to a given passage or the entire musical work from which the passage is taken. In this way, affective qualities can be attributed to the musical content, for example, frustration, relief, anger, pleasure, and so on.

Next, referential scenarios can be constructed. For instance, one might ask whether a particular musical piece is suitable as the background music in a cartoon where Tom the cat is trying to catch Jerry the mouse. Finally, one can provide descriptions of the perceived musical experience that are based on our immanent knowledge and personal experience. This is illustrated in the diagram below.



Another innovation in this project on image schemas is the attempt to formalize these levels of musical interpretation. For example, let us consider the point of highest musical energy in the given Beethoven excerpt, and how we even arrived at the interpretation of the dynamic change in musical energy that occurs during the perception of the piece. One possible approach is to examine formal parameters inherent to individual musical notes and assign positive and negative values corresponding to increased or decreased levels of intensity. Thus, if we take into account the number of notes in a chord, the duration of the notes, interval leaps, volume, and so on, as individual factors, each contributing one point to the increase in dynamics, we obtain energy values, of which the highest, in the last three previously presented diagrams, occurs near the end of the musical notation. In other words, there is also a mathematical support for the perception of variable energy observed in the piece. This is illustrated in the diagram below.



In the same way, we can attempt to formalize the schematic structure beyond the purely perceptual level. For example, in addition to the overall increase and decrease in energy levels, we can track the onset and change in intensity of at least five image schemas: path, force, balance, link, and containment. Thus, in the given excerpt, we can see that the (vertical) path is most pronounced in the F minor chord, as it includes a leap from E4 in the bass clef to C6 in the violin clef, i.e., more than an octave and a half, which in our notation can be marked as <P++>. In addition, the fortissimo on this chord increases the intensity/force, which we can represent as <F++>, while the balance is positive (in our notation <B+>), since we have a tonic chord, even though it is not particularly intense or forceful. Overall, the strongest presence of schemas is again found in the structurally most important chord, providing the climax of the segment.

During the project in which this brochure was produced, special attention was devoted precisely to such formalizations in large linguistic and musical corpora.

5. Image schemas and pedagogical work

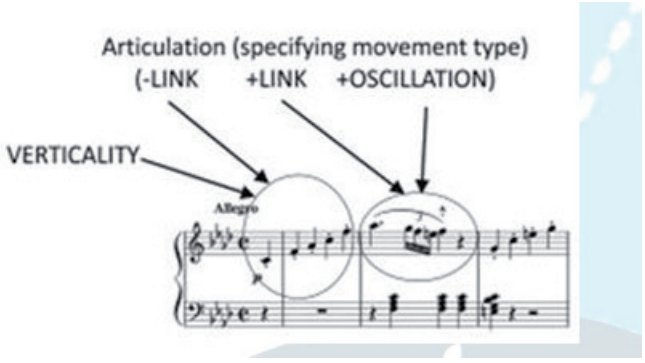
Thus, we arrive at the consideration of whether the aforementioned viewpoints and concepts can also be used for pedagogical purposes. Our opinion is that not only can they be used, but that referencing them can significantly enhance teaching. We will attempt to support this position with some practical suggestions.

For example, within the context of children's music education, real-world scenarios can be creatively developed to naturally introduce how musical structure unfolds. A teacher could first creatively prepare children to understand and use the described concepts. This could be done by asking children to talk, for instance, about balance. During this, no musical background would be used. For example, the

teacher could ask what balance means, how relationships of balance can be applied using the example of a jeweler's scale or another (analogous) scale, what is required for them to maintain balance if standing on one leg, how balance is lost, and what happens when someone truly loses balance – and ask the children to demonstrate all of this using their own example.

Additionally, the discussion could focus on different ways of moving along a path – for example, walking slowly, running, hopping continuously without pauses, or hopping with pauses between jumps. The children could then be asked to demonstrate these various forms of movement themselves. Similarly, the conversation with the children could first explore how they would express a strong force, for instance, when needing to move a heavy object or when they feel angry.

In a second step, the children could be asked to verbalize scenarios that incorporate the relevant image schemas. As in the first step, there would still be no background music. However, if the first melody they would later listen to included, at the beginning, a transition from lower to higher notes played in a detached or interrupted manner, before reaching the first peak of melodic intensity that prevents equilibrium, followed by oscillation (as is the case with the music illustrated in the diagram below), the task could be for the children to imagine a scenario that includes such scenes.



Of course, the children would provide different narratives during this activity. For example, one child might tell a story about a cat climbing stairs quickly and in a choppy manner to catch a mouse, only for the mouse to escape by wriggling out of the cat's paws, which would correspond to oscillatory movement. Immediately after this, the third step would follow, which involves playing the musical excerpt. In the fourth step, the children would be asked to identify the moments in which a particular image-schema concept appears. This would be done with verbal support from the teacher – not in the sense of giving instructions on identifying examples of specific image schemas, but, for instance, by asking them to raise their hand when they hear the cat trying to catch the mouse or when the mouse wriggles free from the cat's paws, and so on. Finally, in the fifth step, the children would be asked to perform the story they created from the music they heard and verbally processed, to the best of their abilities.

This is, of course, a fully embodied activity in which children create a kind of bodily notation for the music they are exposed to. The main advantage of this approach is the early combination of music and a form of dance, pedagogically oriented toward understanding relevant musical concepts in a fun and friendly, yet structured, environment. Moreover, this approach is assumed to work well in inclusive music classrooms, as children with visual impairments could participate fully in such activities. In fact, visually impaired children could make this activity even more engaging because their understanding of musical movement does not have to be horizontal-vertical; it could, for example, follow the direction of a clock's hands or describe the notes they hear as "large" and "small," as shown in some studies (e.g., Antović, 2009). Similarly, children with hearing impairments could benefit from an embodied response to music they cannot physically hear, or hear less clearly, by following the movements of their peers and immersing themselves in the musical structure through their own bodily movements. An additional benefit of this approach in music education, especially for children, would be the use of techniques for free conversation and reflection, aimed at enhancing children's imaginative and narrative skills.

The concepts presented above could also be applied in (foreign) language teaching. For example, in the case of the orientation image schema *up–down*, it could be applied in foreign language lessons in the following way: first, vocabulary belonging to the relevant semantic field would be selected. In the case of English, this could include lexemes such as *upper, top, bottom, hill, take off, plummet, soar, rise, fall, set, ground, sky, ascend, descend*, etc. Next, exercises could be created in which students are asked to insert the appropriate form of a given lexeme (provided in parentheses) into a sentence.

It could look like this:

Task 1: Literal Sense

1. The airplane ___ into the sky. (take off)
2. She climbed the ___ of the mountain. (top)
3. The ship sank to the ___ of the ocean. (bottom)
4. He placed the vase on the ___ shelf. (upper)
5. The balloon ___ into the air. (rise)
6. The rocket ___ into space. (launch)
7. The eagle ___ high above the mountains. (soar)
8. The submarine sank to the ___ of the ocean. (depth)
9. The rocket ___ into the atmosphere. (ascend)
10. The sun ___ behind the mountains. (set)

Task 2: Metaphorical Sense

1. Her career really ___ after she graduated. (take off)
2. He's been feeling on ___ of the world lately. (top)
3. After the news, she felt at the ___ of despair. (bottom)
4. His spirits were in the ___ region. (upper)
5. Her mood ___ after receiving the good news. (rise)
6. His spirits ___ when he heard the news. (launch)
7. Her confidence ___ after receiving praise. (soar)
8. She felt ___ of despair after the loss. (depth)
9. The company's success ___ to new heights. (ascend)
10. His mood ___ after the challenging day. (set)

This approach is similar to the one used in Wright (1999), albeit with different types of metaphors rather than orientation-based ones. Elements of this way of organizing didactic material are also present in the phrasal verb dictionary *Macmillan Phrasal Verbs Plus* (published in 2005). One of the most significant aspects of the organization of this dictionary is the special sections for various phrasal particles, including particles such as *down* and *up*. In these sections, the development of different figurative meanings of such particles from their more basic meanings is presented, which we have already touched upon above, along with accompanying exercises.

6. Additional research in the field of language teaching

Using the associative method with students of Serbian Studies and English Studies at the Faculty of Philosophy in Niš, Aleksandra Janić and Marta Veličković (2023a, 2023b, 2023c) focused on recent noun and adjective Anglicisms excerpted from the *Serbian Dictionary of Recent Anglicisms* (2021) on the one hand, and their established Serbian counterparts on the other. In other words, the recent Anglicisms and their established counterparts were presented to philologically oriented participants as stimuli, and the lexeme-responses obtained through associative methods were analyzed. The results obtained are significant for teachers of both English and Serbian, as native and foreign languages.

Regarding recent noun Anglicisms and their counterparts (2023a; 2023c), the corpus included 40 pairs of examples such as *browser/претраживач*, *office/канцеларија*, *party/журка*, *popcorn/кокице*, *reseller/препродавац*. After analyzing the types of associative responses to recent noun Anglicisms compared with their Serbian counterparts (2023a), the following Anglicisms emerged as the most acceptable: *popcorn, gift, file, jackpot, cash, sticker*. It was shown that native Serbian speakers provide responses of a paradigmatic type to recent noun Anglicisms as stimuli. On the other hand, the number of activated semantic frames and/or idealized cognitive models is smaller for recent Anglicisms than for their established Serbian counterparts. Strong tendencies were observed for the full integration of the analyzed recent noun Anglicisms into the lexical system of the Serbian language. Namely, considering the influence of foreign culture, the analyzed recent Anglicisms will, through various connotations, contribute to the mental lexicon of Serbian speakers. Therefore, the acceptability of recent Anglicisms should not be reduced solely to a criterion of necessity but should be represented on a scale.

Based on the analysis of lexeme-responses in the form of synonyms, hyponyms, or hypernyms relative to the recent Anglicism as a stimulus, it was observed that the meaning of recent noun Anglicisms and their established counterparts provided in the *Serbian Dictionary of Recent Anglicisms* (2021) is not always perceived as completely synonymous; there is room for meaning specification. It is also noteworthy that linguocultural elements were observed for 66.25% of the lexeme-stimuli, with foreign cultural influence present in Anglicisms and domestic cultural

influence in their established Serbian counterparts as stimuli.

When comparing the types of lexeme-responses between Serbian Studies students and English Studies students, the following tendencies emerged: 1) English Studies students more frequently produced hapax synonymous responses to recent Anglicisms as stimuli, which aligns with their higher level of English proficiency; 2) Serbian Studies students were more inclined to provide general responses, i.e., hypernyms relative to the given stimulus, whereas English Studies students tended toward hyponyms; 3) the responses of English Studies students were more often encyclopedic in nature compared to those of Serbian Studies students, leading to the conclusion that the level of English proficiency influences both the degree to which Anglicisms are accepted and how they are interpreted and understood.

Regarding recent adjective Anglicisms and their Serbian counterparts (2023b), the corpus included examples such as *асистиран/потпомогнут*, *изгоинг/лежеран*, *кјут/симпатичан*, *промптан/брз*, *релакс/опуштајући*. From the main conclusions, the following points are highlighted: 1) the most frequent responses consisted of nouns forming a syntagma with the stimulus, followed by lexemes connected to the stimulus in terms of encyclopedic knowledge, and then those related through synonymy with the stimulus; 2) although all analyzed recent Anglicisms had an established equivalent in Serbian, they were not equally familiar to the participants nor equally acceptable; 3) recent Anglicisms that, in the study, showed a high level of familiarity/acceptability are characterized by synonymy/near-synonymy or by specific usage patterns that introduce new meanings into the Serbian language.

As expected, there were more instances of missing responses when the stimulus was a recent Anglicism than when it was its established Serbian counterpart, indicating a low level of familiarity/acceptability for the analyzed recent adjective Anglicisms. Regarding the dominant part of speech among the responses, nouns predominated, followed by adjectives. It is precisely in the syntagmatic combinations of the analyzed adjectives and the lexeme-responses in the form of nouns that the contextually specific meanings of the Anglicisms become visible. Additionally, usage patterns were more clearly defined for established Serbian counterparts than for the recent Anglicisms. Serbian language majors tended to establish syntagmatic connections with the lexeme-stimuli through their associative responses, whether the stimuli were recent Anglicisms or their counterparts, whereas English Studies students, in response to Anglicisms as stimuli, more frequently produced synonyms, nouns forming syntagmatic combinations with the stimulus, hyponyms, and antonyms.

7. Image schemas and public discourse

Finally, we would like to briefly draw attention to the use of image schemas in public discourse, for example in politicians' speeches, as part of a broader, typically unconscious reliance on them in everyday speech and writing, as well as in various types of language use.

For example, whenever people talk about whether a country is or is not "on the right path," whether it is or is not on the road toward the EU, or whether a certain party or coalition will move toward the goal of achieving a decisive electoral victory over political opponents by a certain point in time (e.g., the start of a new year, upcoming elections), the *path* schema emerges.

Similarly, when one says, for instance:

- that a smaller or larger part of the world is characterized as a region of poverty and misery, or a region marked by a large outflow or influx of population;
- that it appears a leader is in an invisible cage accessible only to the most loyal, which might attempt to explain someone's lack of understanding of the difficulties faced by ordinary people;
- that a party will enter parliament or leave the government, and so on;

one can speak of the presence of the *containment* schema.

When someone talks about:

- having difficulty defeating their political opponents, or perhaps losing to them by a narrow margin;
- that it requires great effort to bring about a change in public opinion on a certain issue;
- that a presidential candidate has demolished someone's argument in a live televised debate, and so on;

the *force* schema appears.

Whenever people talk about:

- the need to achieve a balance in politics between what is necessary and what is possible;
- a disruption in the balance between the executive, legislative, and judicial powers, and similar situations;

we can also speak of the *balance* schema.

All in all, image schemas, as a preconceptual foundation of language (but, as we have seen, not only language), are in fact ubiquitous. It is up to each individual to determine for themselves how meaningful, justified, and purposeful their use is − especially in the realm of politics − for a better understanding of the speaker's message, and to what extent it may contain elements of manipulation.

References

Antović, M., Jovanović, V. Ž., & Popović, M. (2024) From spatial perception to referential meaning: convergent image schemas in the music of and texts about Beethoven's piano sonatas. *Frontiers in Psychology*, *15*, 1-15.

Antović, M. (2024). *Image Schemas in Musical Structure − Towards a New Tool for Music Instruction for Young Participants*. Conference presentation from the 12th triennial

conference of the European Society for the Cognitive Sciences of Music, 03-06 July 2024, York, La Plata, Melbourne (https://sites.google.ćom/york.ać.uk/esćom12/home).

Antović, M. (2022). *Multilevel Grounding: A Theory of Musical Meaning*. Abingdon & New York: Routledge.

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

Janić, A., & Velićković, M. (2023a). The association networks of select recent nominal Anglicisms and their Serbian language equivalents. *Vestnik of Saint Petersburg University. Language and Literature, 20*(4), 888–905.

Janić, A., & Velićković, M. (2023b). Recent adjectival Anglicisms and their Serbian equivalents: an associative approach, *Зборник Матице српске за филологију и лингвистику, 66*(1), 93–118.

Pavlović, V., & Janić Mitić, A., Mitić, I. (2024). Motion-related image schemas in Serbian journalistic articles: a corpus-based study. *Review of Cognitive Linguistics: Online-First Articles*. Published online: 3 December 2024. DOI https://doi.org/10.1075/rćl.00210.pav.

Rasulić, K. (2004). *Jezik i prostorno iskustvo: konceptualizacija vertikalne dimenzije u engleskom i srpskohrvatskom jeziku*. Beograd: Filološki fakultet Univerziteta u Beogradu.

Rundell, M. (Editor-in-Chief) (2005). *Macmillan Phrasal Verbs Plus*. Oxford: Macmillan Education.

Veličković, M., & Janić, A. (2023). An Analysis of the Associative Networks of Recent Nominal Anglicisms of Serbian and English Language Majors, *Folia Linguistica et Litteraria, 45*, 43–63.

Wright, J. (1999). *Idioms Organiser: Organised by Metaphor, Topic, and Key Word*. Hove: Language Teaching Publications.

A TWO-STAGE MACHINE LEARNING SYSTEM FOR THE ANNOTATION OF VISUAL SCHEMAS: MODELS FOR BOUNDARY DETECTION AND MULTI-CLASS CLASSIFICATION OF VISUAL SCHEMAS

Science Fund of the Republic of Serbia
Programme: IDEAS
Project: Structuring Concept Generation with the Help of Metaphor, Analogy, and Schematicity (Project No. 7715934)
Acronym: *SCHEMAS*
Prepared by: Mladen Popović, MSc

1. Introduction

Over the past several decades, cognitive linguistics has highlighted the complex interrelationship between conceptual structures and linguistic expressions, demonstrating that language is not an autonomous formal system but is instead deeply rooted in embodied cognition (Evans & Green, 2006; Langacker, 2008). Since the early work of Lakoff, Johnson, and their contemporaries, it has become a foundational principle that meaning is grounded in sensorimotor experience and shaped by patterns of bodily interaction with the environment (Johnson, 1987; Lakoff, 1987). Within this theoretical framework, image schemas have emerged as a particularly important concept: they represent recurring patterns of perception and movement that influence the structuring of more complex conceptualizations (Lakoff, 1987: 28–29). Image schemas are conceptualized as basic, pre-linguistic gestalts arising from embodied experience and recurring across multiple conceptual domains, providing a framework upon which more complex metaphorical and abstract reasoning is constructed (Lakoff, 1987: 29–30).

Although their role in conceptual meaning and metaphorical extension has been extensively studied, comparatively less attention has been devoted to how contemporary computational approaches can model the occurrence of image schemas in texts. This gap not only raises theoretical questions but also poses a practical challenge for computational approaches to language: how can we design models that automatically detect these image-schematic patterns in natural texts? Addressing such questions opens up a fertile field of inquiry. If image schemas are indeed foundational to our conceptual apparatus, then examining how machine-

learning models learn to associate linguistic sequences with particular schemas may provide deeper insight into how language is constructed, processed, and understood.

At the core of many studies in cognitive linguistics lies the notion of image schemas: dynamic, embodied patterns of experience that arise through repeated sensorimotor interactions with the environment (Johnson, 1987; Lakoff, 1987; Mandler, 2004). Image schemas are not static mental images; rather, they constitute continuous and interactive patterns of experience, such as movement along a path, the application of force, or being located within a bounded space (Johnson, 1987; Lakoff, 1987; Mandler, 2004). According to Johnson (1987), these schemas emerge early in cognitive development, as infants learn to maintain bodily balance, manipulate objects, notice objects entering and exiting containers, and orient themselves in space.

Because image schemas are rooted in bodily experience, they are commonly regarded as universal or near-universal cognitive structures. They arise from embodied interactions shared by all humans, such as the sense of vertical orientation (UP–DOWN) or the experience of boundaries and enclosed spaces (CONTAINMENT) (Johnson, 1987; Evans & Green, 2006). However, the specific linguistic and cultural elaborations of these schemas may vary, allowing a certain degree of universality at the conceptual level to coexist with linguistic diversity.

Despite this rich body of research, scholarly attention to image schemas has largely focused on their conceptual and semantic dimensions. Cognitive linguists have examined in detail the ways in which image schemas shape meaning, conceptual metaphors, and reasoning (Lakoff, 1987; Johnson, 1987), but the question of how these conceptual patterns manifest at the linguistic level − how they are encoded, cued, or reinforced by linguistic structures − has remained relatively underexplored.

Computational linguistics has, in most cases, approached the problem of semantics in natural language from the perspective of word sense disambiguation (using vector-based or Bayesian methods), the establishment of textual entailment, as well as tasks such as segmentation, summarization, and named entity recognition (Dahlgreen, 1998; Helbig, 2005; Kapetanios et al., 2013). Although some early studies pointed to possible correlations between syntactic constructions and fundamental image schemas (Clausner & Croft, 1999; Wachowiak & Gromann, 2022), the field as a whole has not yet fully confronted the challenge of mapping image schemas onto patterns of natural language.

Moreover, the machine-learning perspective on this problem – the development of algorithms for the detection and annotation of potential image-schematic structures in text − remains largely unexplored. Methods proposed in previous research range from hand-crafted, corpus-based algorithms to unsupervised machine-learning approaches that use part-of-speech (POS) tags to identify combinations of nouns, verbs, and prepositions indicative of image schemas (Dodge & Lakoff, 2005; Bennett & Cialone, 2014; Gromann & Hedblom, 2017; Wachowiak, 2020; Wachowiak & Gromann, 2022). The most recent of these approaches employs a variant of BERT (Bidirectional Encoder Representations from Transformers), a language model that represents texts as sequences of vectors using a transformer architecture and attention

mechanisms (including self-attention), which has been further fine-tuned to classify image schemas using entire text sequences as input (Vaswani et al., 2017; Devlin et al., 2019; Wachowiak & Gromann, 2022). However, it should be noted that none of the approaches described here attempts to model the presence of multiple schemas within a given linguistic sequence, nor do they decompose larger structures, such as sentences, into smaller schema-bearing segments.

The central research question of the present study concerns how image schemas, as conceptual building blocks, can be made machine-readable and learnable, such that a computational model can process and identify schemas occurring in any linguistic segment, regardless of its length. If image schemas indeed constitute the foundation of a wide range of conceptual and linguistic phenomena, it is reasonable to assume that their influence can be detected not only at the semantic or conceptual level, but also at the level of individual words and their combinations.

Wachowiak and Gromann (2022), in their discussion of possible models for the annotation of image schemas, emphasize that an adequate image-schema classifier should interact directly with individual words (or their combinations), such that the model's output simultaneously *locates* and *annotates* schematic complexes.

It is also worth noting that, in their discussion of multi-label annotation and its application to image schemas, these authors identify a particular challenge − specifically, the problem of *input scope*. By way of illustration, consider the input to their classifier, which takes an entire sentence as input and returns only a single label. If we assume that multiple image schemas can occur within a single sentence, the classifier would first need to segment the sentence into smaller units and then predict schemas with respect to those individual segments, especially in cases where a schema is realized across multiple words.

To see why this is not a trivial problem, we may consider an example from Wachowiak and Gromann's model (2022), in which the words *beyond*, *attractive*, *answer*, and *white* are labeled as indicators of the CONTAINMENT schema. This appears to be a side effect of the model's failure to clearly distinguish between lexical elements that contribute to the realization of a schema and those that do not. Addressing this issue requires overcoming yet another obstacle: the relative lack of empirical research employing large-scale corpus-based methods to investigate how image schemas are realized in natural language.

Most early work in cognitive linguistics on image schemas relies on introspective data, small sets of examples, or carefully constructed experimental materials (Gibbs et al., 1994; Gibbs & Colston, 1995; Dodge & Lakoff, 2005; Bennett & Cialone, 2014). Consequently, there are no publicly available corpora specifically designed for training schema-classification models, which forces researchers to create suitable datasets *ad hoc*. As a result, training data are limited in size and are most often computationally derived from texts (e.g. Gromann & Hedblom, 2017; Wachowiak & Gromann, 2022). If a hypothetical image-schema classifier is to perform at a satisfactory level, it will require both large-scale data and an efficient processing logic capable of highlighting the relevant elements within a sentence.

To address these questions, the present study draws on the SCHEMAS corpus – a purpose-built dataset compiled for the investigation of image schemas and their linguistic realizations. The SCHEMAS corpus was created through the manual annotation of image schemas in texts drawn from various online newspapers. It consists of two smaller subcorpora, one in Serbian and one in English, and was originally used to examine the distribution of image schemas across different languages. In the present context, this carefully prepared corpus enables us to train (1) a processing model that identifies meaning-bearing segments of text (i.e. textual "chunks") in which schemas occur, as well as (2) a multi-label classification model trained on naturalistic data. The decision to use an already annotated corpus is primarily motivated by the fact that it allows us to avoid issues inherent to synthetic data, while at the same time providing sufficiently diverse and realistic input that reduces the risk of overfitting.

Crucially, both models must be context-sensitive in order to handle the simultaneous occurrence of multiple schemas effectively. Consider, for example, the sequence "*He was falling into a deep depression.*" An annotator following the SCHEMAS project guidelines might annotate the entire sequence with schema labels, placing the tags <PATH><FORCE><CONTAINMENT> after the word "*depression.*" It is worth noting that the primary contributions to the <PATH> and <CONTAINMENT> schemas are found in the segment "*was falling ... into a deep depression*," whereas the main contribution to the <FORCE> schema is localized in the word "*falling.*"

Thus, for a model to accurately predict the occurrence of these schemas, it must be capable of contextual understanding. To capture context, both the segmentation model and the classifier model were built using BERT, a deep neural language model based on the now widely adopted transformer architecture (Vaswani et al., 2017; Devlin et al., 2019). BERT processes context using the transformer's self-attention mechanism in a bidirectional manner. This means that, unlike models such as GPT, BERT processes tokens in both directions simultaneously, so that word representations during training are learned by taking into account tokens to both the left and the right of the target word.

This is achieved through the transformer's multi-head self-attention mechanism, which allows each token to attend to every other token by computing attention weights that indicate the relevance of each token with respect to the token being processed. The architecture is integrated during the training process, in which BERT randomly masks input tokens with the goal of later predicting the masked words. Because prediction is performed by taking into account tokens both to the left and to the right of the target words, the model is able to learn deep contextual interactions from all directions. After training is completed, the representation of each token also encodes information about all the other tokens in the input sequence. These representations (embeddings) are subsequently fine-tuned for specific tasks such as identifying schema-bearing segments and annotating schemas, as is the case in the present study. This is accomplished using a feed-forward classifier (often referred to as a dense or linear layer), a small neural network module placed on top

of BERT's final output, which transforms its high-dimensional contextual encodings into task-specific predictions (Devlin et al., 2019).

Particularly important for our task is the fact that BERT, as a result of its training objective involving masked word prediction within a sentence, acquires a rich, contextualized representation of language. Moreover, once the model has been pretrained, it can be further adapted to a specific task by adding an additional layer. Finally, the model is optimized in an end-to-end manner, meaning that no additional feature engineering is required; instead, the model autonomously learns to attend to the features most relevant for the given task. Given that our corpus contains annotations that correspond to more or less clearly defined text sequences, the attention mechanism is considered a particularly important component of the overall system (Wachowiak & Gromann, 2022).

In the remainder of the paper, this study will first attempt to model the procedure followed by a human annotator, in order to highlight some of the requirements that the computational model should satisfy. It will then describe the process of preparing the corpus for model training and testing, as well as the overall structure of the solution (the processing pipeline). The final part of the study will demonstrate how the two models operate in synergy and will propose possible improvements. One of the main drivers of the overall model architecture was the set of observations presented in Wachowiak and Gromann (2022); accordingly, we will occasionally refer back to their solution in what follows, as it represented the closest available analogue to our own at the time of writing.

2. The method

Let us now consider the task of the human annotator. Upon encountering a text, a hypothetical annotator, following the SCHEMAS project annotation guidelines (as outlined in Antović et al., 2023; Figar & Veličković, 2023), might read the text and, taking into account the examples provided in the guidelines, annotate a sample sentence as follows:

> [...] "This legislation is a giant step forward<ms><F><P++><spec><forward> in our fight to combat<ms><F+><L> the fentanyl crisis, crack down on the dealers peddling<ms><F><P+><L> death in our communities, and accelerate<ms><F+><P+><spec><forward><L> our state's public health response to get this deadly drug off our streets<s><F+><P><L--> and save lives," House Speaker Alec Garnett, a Democrat, said after the bill's passage<ms><F><P+><spec><end path>." [...]

Sequences beginning with <ms> or <s> and ending with the final symbol (>) represent examples of schematic complexes. In this case, they indicate that a

particular stretch of text has been recognized as containing schemas − specifically, <FORCE>, <PATH>, <LINK>, <BALANCE>, <CONTAINMENT> − as well as a scalar modification (a <SCALE> schema that "applies" to these five core schemas). Scalar modifications in the annotation set are marked with either a plus or minus sign, indicating whether the valence is positive or negative, with multiple repeated signs denoting greater intensity (for example, a <FORCE> schema with one plus sign is interpreted as more intense than the same schema without a sign).

The presence of multiple annotation clusters (conceptual combinations composed of several image schemas) within a sentence indicates that the annotator has associated specific words from the sentence with particular schematic clusters. For example, the sentence "*This legislation is a giant step forward*" is annotated as <ms><F><P++><spec><forward>, whereas "*in our fight to combat*" is annotated as <ms><F+><L>. When text is annotated in this manner, it becomes possible to conduct further analyses to determine both absolute and relative frequencies of schemas, either as clusters or as individual schemas.

It should be noted, however, that a single sentence can contain multiple clusters, while no cluster in the corpus receives information from elements outside the boundaries of that sentence. It is possible that annotators, while reading the final segments of one sentence, carried over information into the next, resulting in a "bleed-over" of information between structures that, in theory, should constitute separate meaning units. Nevertheless, for the purposes of model training, it was decided that sentence boundaries were to be retained as "hard" constraints, based on a preliminary analysis of segmentation (chunking) logic, which showed that this approach successfully preserves semantically significant segments in all 300 randomly selected examples from the corpus. Accordingly, a potential annotation procedure can be represented approximately as follows:

1. The annotator familiarizes themselves with the protocol and criteria for identifying each image schema.
2. The annotator keeps these criteria "active" in memory while reading and processing the incoming text.
3. During reading, the processing of specific lexical elements and their contextual frames increases the likelihood that certain schemas will be detected and annotated.
4. When this cumulative probability reaches a certain threshold, the schemas are recognized and recorded in the annotation. This accumulation occurs simultaneously for multiple schemas.
5. The moment of annotation interrupts the existing accumulation of probability, meaning that after annotating a particular segment within a sentence, the process restarts from the beginning.
6. Since no schemas extend beyond sentence boundaries, only local sequences within the sentence itself are relevant for accumulation.

Let us consider how a computational variant of the annotator could emulate a human annotator. At the most basic level, the system must process each input text

and segment it into sentences. Each sentence is then further divided into individual tokens, allowing an attention-like mechanism to track interactions between tokens and gradually accumulate probabilities for the presence of specific schemas within a segment. Once these probabilities exceed a certain threshold, the computational annotator would finalize the chunk and assign the relevant schema labels to that segment, in a manner analogous to how a human annotator stops reading, records the recognized schemas, and then continues. With this in mind, the first step in developing our predictive model consists of formalizing a method that reliably "captures" the portion of text located to the left of a given schematic cluster.

To train a BERT-based model, the corpus must be segmented so that each schematic cluster is clearly associated with its direct "source," i.e., the text contributing to its detection. To this end, we used the NLTK library to create an algorithm that searches the text to the left of a cluster and stops when it encounters a sentence boundary or another schematic cluster (Bird, 2006). Sentence segments contributing to a given cluster are referred to as "segments" or "chunks," as they often include extraneous information. These segments are further processed by assigning labels to the tokens: the onset token of each segment receives a specific label, while all the other tokens within the segment are assigned interior token labels. Tokens that do not contribute to schema recognition are labeled as empty tokens. These labels are then fed into a modified version of BERT, based on the "BERT base uncased" model available on the Hugging Face platform (Wolf et al., 2020). The task of identifying these segments is, by its nature, similar to other sequence classification tasks, for which BERT has already demonstrated reliable performance (Wolf et al., 2020).

Because chunks can exhibit significant syntactic variation − including fully formed syntactic structures and extraneous fragments (e.g., "the fentanyl crisis, crack down on the dealers peddling<ms><F><P+><L>") − the computational annotator must also "learn" how to selectively evaluate or ignore portions of each chunk. In other words, the system cannot rely solely on a single deterministic algorithm; rather, it must adapt to different linguistic configurations through dynamic adjustment of the probabilities associated with encountering particular schemas. This adaptive capacity is precisely what makes a machine learning approach indispensable, as it allows the model to generalize from annotated examples and handle the inherent variability of language in natural contexts. Similar to the task of identifying schema-bearing chunks, a BERT model was again chosen as the foundation, trained on textual chunks obtained using the previously described processing algorithm. Each schematic cluster associated with a chunk was also extracted and encoded as a "one-hot" vector (Brownlee, 2020).

All of the code was implemented in Python, and the corresponding Jupyter notebook is available at the following address:

https://github.com/MladenPopovicFilFak/ProjectSchemasAnnotator.

The segmentation model and the annotation model can be found here:

https://huggingface.co/MladenIgnatum/Segmentation_Model https://huggingface.co/MladenIgnatum/Annotation_Model.

3. Text segmentation tool – From corpus to training sets

The first step in constructing our synthetic annotator involves preprocessing the annotated text so that the resulting dataset preserves meaningful structures (i.e., chunks and their corresponding annotations) while remaining machine-readable. At the most basic level, the parsing algorithm for the human-annotated corpus must:

1. Preserve the portion of the text located to the left of the annotation, and
2. Preserve the annotation itself.

This directly informs the architecture of the annotation model. Considering the model requirements to both segment text and perform annotation, it was decided that the process should proceed in two phases. First, a submodel based on BERT is trained on the first data segment (1) to learn to identify chunks containing schemas and extract them from any input text. Then, a second submodel, also based on BERT, is trained on the textual segments from (1), together with their corresponding schematic clusters (2). This second model used the output of the first submodel to predict the presence of schemas. The code and instructions are detailed below.

Cell 1 handles the import of the necessary libraries. **chardet** is used for encoding detection, **nltk** for linguistic parsing, while **torch** and **scikitlearn** are used for managing the learning environment. The **transformers** library is employed for working with BERT and related operations (Chardet n.d.; Collobert et al., 2002; Bird, 2006; Pedregosa et al., 2011; Duchesnay, 2011).

Boundary labels are used to annotate segments carrying schemas. The first token of a segment is labeled the B-Chunk (indicating the beginning of a given segment), I-Chunk labels denote words between the start of the segment and the associated schematic cluster, and the O label applies to tokens that do not contain schemas.

```
################################################
CELL 1: SETUP, INSTALLS, AND IMPORTS
################################################

from google.colab import drive
drive.mount('/content/drive')

import os
import re
```

```
import chardet
import nltk
import torch
import numpy as np
from nltk import sent_tokenize
from sklearn.model_selection import train_test_split
from sklearn.metrics import f1_score

from transformers import (
  AutoTokenizer,
  AutoModelForTokenClassification,
  AutoModelForSequenceClassification,
  Trainer,
  TrainingArguments
)

# Download NLTK sentence tokenizer data
nltk.download('punkt')
nltk.download('punkt_tab')

# Global configs
MODEL_NAME = "bert-base-uncased"   # Or another HF model
MAX_LEN = 128

# For Stage 1 boundary detection
BOUNDARY_LABELS = ["O", "B-CHUNK", "I-CHUNK"]
label2id_boundary = {lab: i for i, lab in enumerate(BOUNDARY_LABELS)}
id2label_boundary = {i: lab for i, lab in enumerate(BOUNDARY_LABELS)}

# For Stage 2 multi-label classification
SCHEMA_LABELS = ["P", "F", "L", "B", "C"]
SCHEMA2ID = {s: i for i, s in enumerate(SCHEMA_LABELS)}
```

Cell 2 is responsible for parsing the raw textual data to identify and extract meaningful segments based on the inline annotations. This parsing logic is crucial for preparing the data for both phases of the procedure − **phase 1** (boundary detection) and **phase 2** (schema classification). Specifically, **cell 2**:

- **Splits the text into segments:** Uses regular expressions to identify and separate parts of the text marked with inline annotations (e.g., <P>, <F+>).
- **Processes the annotations:** Simplifies complex labels into basic schema labels, ensuring consistency and relevance.
- **Prepares data for training:** Structures the data into formats suitable for boundary detection and schema classification tasks.

```
#####################################################
CELL 2: PARSING LOGIC FOR INLINE TAGS -> CHUNKS
#####################################################

# Define SCHEMA_LABELS and SCHEMA2ID (Ensure consistency)
SCHEMA_LABELS = ["P", "F", "L", "B", "C"]
SCHEMA2ID = {s: i for i, s in enumerate(SCHEMA_LABELS)}
SCHEMA_ID2LABEL = {i: s for i, s in enumerate(SCHEMA_LABELS)}

def parse_clusters_in_sentence(sentence):
  """
  Splits a sentence on consecutive <...> tags.
  Associates each tag group with the preceding chunk.
  Returns a list of dicts:
   [
    {
     "chunk_text": "...",
     "raw_tags": [...],
    },
    ...
   ]
  """
  # Pattern to identify consecutive tags as one group
  pattern = r'((?:<[^>]+>)+)'
  parts = re.split(pattern, sentence)

  chunks = []

  # Iterate over parts in pairs: text + tags
  for i in range(0, len(parts), 2):
    text_chunk = parts[i].strip()
    tags = []
```

```
        if i + 1 < len(parts):
            tags = re.findall(r'<([^>]+)>', parts[i + 1])
        if text_chunk:
            chunks.append({
                'chunk_text': text_chunk,
                'raw_tags': tags.copy()
            })

    return chunks

def collapse_raw_tags(raw_tags):
    """
    Convert tags like ["ms", "F+", "P++"] to a set of [P, F, L, B, C].
    Discard irrelevant tags (ms, spec, forward, etc.).
    """
    core_set = set()
    for rt in raw_tags:
        if not rt:
            continue
        base = rt[0].upper()  # Ensure case insensitivity
        if base in {"P", "F", "L", "B", "C"}:
            core_set.add(base)
    return core_set

def parse_text_into_stage_data(raw_text, tokenizer):
    """
    1) Sentence-splits the text.
    2) For each sentence, parse chunk boundaries (Stage 2) + create
B/I/O (Stage 1).
    Returns:
      stage1_data: [{"tokens": [...], "labels": ["B-CHUNK","I-CHUNK",...]}]
      stage2_data: [{"chunk": "...", "labels": [0/1,...]}]
    """
    sentences = sent_tokenize(raw_text)
    stage1_data = []
    stage2_data = []

    for sent_idx, sent in enumerate(sentences, 1):
        # Identify chunk boundaries
        sent_chunks = parse_clusters_in_sentence(sent)
        print(f"\nProcessing Sentence {sent_idx}: {sent}")
```

```
    print(f"Identified Chunks: {sent_chunks}")

    # Reconstruct clean_sentence without tags
    clean_sentence = re.sub(r'<[^>]+>', '', sent).strip()
    print(f"Clean Sentence: {clean_sentence}")

    # Tokenize the clean_sentence with offset mapping
    encoding = tokenizer(clean_sentence,
return_offsets_mapping=True, add_special_tokens=False)
    tokens = tokenizer.convert_ids_to_tokens(encoding['input_ids'])
    offset_mappings = encoding['offset_mapping']
    print(f"Tokens: {tokens}")
    print(f"Offset Mappings: {offset_mappings}")

    # Initialize labels as "O"
    label_sequence = ["O"] * len(tokens)

    # Track the last assigned character to prevent overlapping
assignments
    last_assigned_char = 0

    for ch_idx, ch in enumerate(sent_chunks, 1):
      chunk_text = ch["chunk_text"]
      raw_tag_set = collapse_raw_tags(ch["raw_tags"])

      if not raw_tag_set:
        # This chunk has no labels, so tokens remain "O"
        print(f"Chunk {ch_idx}: '{chunk_text}' - No Labels Assigned")
        continue  # Do not assign labels to these tokens

      # Find the chunk_text in clean_sentence starting from
last_assigned_char
      start_char = clean_sentence.find(chunk_text, last_assigned_char)
      if start_char == -1:
        print(f"Warning: Chunk '{chunk_text}' not found in
clean_sentence.")
        continue

      end_char = start_char + len(chunk_text)
      print(f"Chunk {ch_idx}: '{chunk_text}' - Start: {start_char}, End:
{end_char}")
```

```
        # Assign labels to tokens within [start_char, end_char)
        first_token = True
        for i, (token_start, token_end) in enumerate(offset_mappings):
            if token_start >= start_char and token_end <= end_char:
                if label_sequence[i] == "O":  # Only assign if not already
labeled
                    if first_token:
                        label_sequence[i] = "B-CHUNK"
                        first_token = False
                    else:
                        label_sequence[i] = "I-CHUNK"

        # Assign Stage 2 labels
        label_vec = [0] * len(SCHEMA_LABELS)
        for t in raw_tag_set:
            if t in SCHEMA2ID:
                idx = SCHEMA2ID[t]
                label_vec[idx] = 1
        print(f"Raw Tags: {ch['raw_tags']}")
        print(f"Assigned Labels: {label_vec}")

        stage2_data.append({
            "chunk": chunk_text,
            "labels": label_vec
        })

        # Update last_assigned_char to end_char to prevent overlapping
        last_assigned_char = end_char

    # Append to Stage1 data
    stage1_data.append({
        "tokens": tokens,
        "labels": label_sequence
    })

  return stage1_data, stage2_data
```

In **cell 2**, the following elements are present: **SCHEMA2ID** and **SCHEMA_ID2LABEL**, dictionaries that map schema labels to unique IDs and vice versa. These mappings are crucial for model training and prediction, as they provide a consistent numerical representation of the labels (Collins & Syme, 1995).

```
# Define SCHEMA_LABELS and SCHEMA2ID (Ensure consistency)
SCHEMA_LABELS = ["P", "F", "L", "B", "C"]
SCHEMA2ID = {s: i for i, s in enumerate(SCHEMA_LABELS)}
SCHEMA_ID2LABEL = {i: s for i, s in enumerate(SCHEMA_LABELS)}
```

The function parse_clusters_in_sentence(sentence) uses regular expressions to detect and group consecutive inline tags within a sentence. The purpose of this function is to identify and group consecutive inline tags in a single sentence and link each group of tags to the preceding textual segment, effectively creating a mapping of text chunks to their corresponding labels. The regular expression pattern r'((?:<[^>]+>)+)' works as follows:

1. Defines a non-capturing group (?:) to exclude the text to the left of the first annotation,
2. Recognizes any sequence of characters within < and > as a single tag <[^>]+>,
3. Matches one or more consecutive tags (?:<[^>]+>)+,
4. Captures the entire sequence of consecutive tags ((?:<[^>]+>)+).

After that, re.split splits the sentence into alternating segments of text and groups of tags. For example, "This is a sample <P>text<F+>." would be split into: ['This is a sample ', '<P><F+>', '.'] (Van Rossum, 2020). The function then iterates through these segments and creates a dictionary containing the text chunks and their raw tags. For example: [{ "chunk_text": "This is a sample", "raw_tags": ["P", "F+"] }]

The function collapse_raw_tags(raw_tags) simplifies the schemas present in the corpus. Considering that the number of possible schemas and their combinations is quite large, the pipeline was initially tested using a reduced set of schemas. All specific tags, such as <forward>, <up>, <down>, etc., were removed, while scalar schemas were consolidated into their base variant: <F+>, <F++>, <F+++>, and other scalar versions were collapsed into a single category <F>.

```
def parse_clusters_in_sentence(sentence):
    """
    Splits a sentence on consecutive <...> tags.
    Associates each tag group with the preceding chunk.
    Returns a list of dicts:
     [
      {
       "chunk_text": "...",
       "raw_tags": [...],
      },
      ...
     ]
    """
    # Pattern to identify consecutive tags as one group
    pattern = r'((?:<[^>]+>)+)'
    parts = re.split(pattern, sentence)

    chunks = []

    # Iterate over parts in pairs: text + tags
    for i in range(0, len(parts), 2):
        text_chunk = parts[i].strip()
        tags = []
        if i + 1 < len(parts):
            tags = re.findall(r'<([^>]+)>', parts[i + 1])
        if text_chunk:
            chunks.append({
                'chunk_text': text_chunk,
                'raw_tags': tags.copy()
            })

    return chunks

def collapse_raw_tags(raw_tags):
    """
    Convert tags like ["ms", "F+", "P++"] to a set of [P, F, L, B, C].
    Discard irrelevant tags (ms, spec, forward, etc.).
    """
    core_set = set()
    for rt in raw_tags:
        if not rt:
```

```
        continue
    base = rt[0].upper()  # Ensure case insensitivity
    if base in {"P", "F", "L", "B", "C"}:
        core_set.add(base)
return core_set
```

The function parse_text_into_stage_data(raw_text, tokenizer) processes raw input texts and creates datasets for training both models. The function outputs two lists of dictionaries: **stage1_data** contains tokens along with their corresponding **boundary labels**, which indicate the start, inside, or outside of a chunk, while **stage2_data** contains the extracted text chunks and their corresponding schemas, represented as multi-hot vectors that indicate the presence or absence of a particular schema (Collins & Syme, 1995). To illustrate, the function returns data of the following form:

```
stage1_data = [
  {
    "tokens": ["Brent", "crude", "'", "s", "rise", "above",
"that", "milestone", "."],
    "labels": ["B-CHUNK", "I-CHUNK", "O", "O", "O", "O",
"O", "O", "O"]
  },
  ...
]

stage2_data = [
  {
    "chunk": "Brent crude's rise above that milestone",
    "labels": [1, 1, 0, 0, 0]  # Example: P and F schemas
present
  },
  ...
]
```

So, if the sentence: “Today another American president faces rising<ms><P><F><Spec><UP> fuel prices, spurred<ms><F+><P++><L+> by a challenge mostly out of his control, an invasion<s><F++><P++><L++><C+> of Ukraine by Russia, a top oil and gas producer intent to use its energy supplies as a weapon when necessary.” is processed by the function, the output of the processing looks like this:

Stage 1 Data: Sentence 1: today: B-CHUNK another: I-CHUNK american: I-CHUNK president: I-CHUNK faces: I-CHUNK rising: I-CHUNK fuel: B-CHUNK prices: I-CHUNK ,: I-CHUNK spurred: I-CHUNK by: B-CHUNK a: I-CHUNK challenge: I-CHUNK mostly: I-CHUNK out: I-CHUNK of: I-CHUNK his: I-CHUNK control: I-CHUNK ,: I-CHUNK an: I-CHUNK invasion: I-CHUNK of: O ukraine: O by: O russia: O ,: O a: O top: O oil: O and: O gas: O producer: O intent: O to: O use: O its: O energy: O supplies: O as: O a: O weapon: O when: O necessary: O .: O

and:

Stage 2 Data:
Chunk 1: Today another American president faces rising
Assigned Labels: ['P', 'F']

Chunk 2: fuel prices, spurred
Assigned Labels: ['P', 'F', 'L']

Chunk 3: by a challenge mostly out of his control, an invasion
Assigned Labels: ['P', 'F', 'L', 'C']

The data from **phase 1** are used for training and validating the segmentation model, while data from **phase 2** are used for training and validating the annotation model. The models are applied in tandem, so that each input text is first segmented using the segmentation model, and then the resulting chunks are passed to the schema-assignment model. This approach allows efficient and precise handling of textual data in two stages, with the goal of ensuring the best possible interpretation and classification of complex schematic structures in the text.

Cell 3 is responsible for transforming unstructured textual data into a structured format, thereby laying the foundation for efficient model training in the following cells. The code in this cell applies previously defined functions and iterates through all files that make up our corpus:

- **Reading raw text files:** It goes through the specified directories to locate and load all .txt files containing raw data.
- **Detection and handling of encoding:** Uses the **chardet** library for accurate detection of each text file's encoding, ensuring correct reading of diverse datasets (**chardet, n.d.**).
- **Parsing text into structured data:**
  - **Phase 1 (Boundary Detection):** Prepares token-level labeled data (BIO — Begin, Inside, Outside) for recognizing chunk boundaries.
  - **Phase 2 (Schema Classification):** Structures data with multi-hot labels corresponding to predefined schema categories for each identified chunk (Wachowiak & Gromann, 2022).

Cell 3 completes the data preprocessing step in the overall procedure. The procedure contained in the first three cells can accept any annotated file as input, provided that the annotation procedure is similar to that used in the original files. The structure of the procedure is as follows: (1) Raw texts with annotations are split into textual chunks associated with specific schema groupings, and (2) for the first model, a dataset is created using boundary labels, while for the second model, a dataset is created by encoding schema groupings as one-hot vectors.

Cell 4 further processes these two datasets so that they can be used as input values for the two models. **Cells 5** and **6** initialize and train these two models: **cell 5** handles the segmentation model, while **cell 6** handles the annotation model. Finally, **cell 7** applies these two models in tandem. In **cell 7**, input text is provided for processing, after which the models jointly perform segmentation and annotation of the input text.

```
##################################################
CELL 3: READ ALL TXT FILES, DETECT ENCODING, PARSE -> STAGE1 &
STAGE2
##################################################
# Define your tokenizer (ensure it matches the one used in parsing logic)
MODEL_NAME = "bert-base-uncased"  # Replace with your specific
model if different
tokenizer = AutoTokenizer.from_pretrained(MODEL_NAME)
FOLDER_PATH = "/content/drive/"  # Change depending on the location
of your input files
all_stage1 = []
```

```
all_stage2 = []
txt_files = [f for f in os.listdir(FOLDER_PATH) if f.endswith('.txt')]
print(f"Found {len(txt_files)} text files.")
for filename in txt_files:
    full_path = os.path.join(FOLDER_PATH, filename)

    # 1) Detect encoding
    with open(full_path, 'rb') as f:
        raw_data = f.read(2048)
        detected = chardet.detect(raw_data)
        encoding = detected['encoding']
        if not encoding:
            encoding = 'utf-8'  # Fallback encoding
            print(f"Encoding not detected for {filename}. Using fallback
encoding 'utf-8'.")

    # 2) Read file with detected encoding
    try:
        with open(full_path, 'r', encoding=encoding, errors='replace') as f:
            file_text = f.read()
    except Exception as e:
        print(f"Error reading {filename} with encoding {encoding}: {e}")
        continue  # Skip to the next file in case of an error

    # 3) Parse text -> stage1, stage2
    stage1_data, stage2_data = parse_text_into_stage_data(file_text,
tokenizer)
    all_stage1.extend(stage1_data)
    all_stage2.extend(stage2_data)

    print(f"Processed file: {filename}")

print(f"\nTotal Stage1 examples: {len(all_stage1)}")
print(f"Total Stage2 examples: {len(all_stage2)}")
```

4. Data for the model and training – segmentation and annotation

The data processed by the algorithm described in cells 1–3 were taken from the SCHEMAS corpus. The specifics of the corpus itself (as described earlier) and the annotation procedure are too numerous to detail here, but one particularly important fact is that each annotator contributed 50,000 words, working in pairs. After completing the annotation, each member of a pair reviewed the annotations of the other member. In the next step of the procedure, another pair (unrelated to the first) additionally verified the annotations of the first pair.

This circumstance significantly influenced the choice of training data – to maintain consistency, it was decided that both models would be trained on a sub-corpus of 100,000 words, composed of the annotations from one pair of annotators. The main assumption was that this choice would preserve a more uniform annotation approach (assuming equal or approximately equal sentence contexts), while simultaneously increasing the size of the training dataset. Furthermore, the number of schemas per category was highly unbalanced, with some being very sparsely represented (e.g., the BALANCE schema), which made it necessary to expand beyond the corpus of a single annotator (50,000 words) to include the rarer schemas. These limitations should be taken into account in subsequent work. **Cells 5**, **6**, and **7** are shown in the appendix because they do not introduce new algorithms.

Cell 4 is responsible for converting the parsed data from **cell 3** into structured datasets suitable for training machine learning models, both in **phase 1** (boundary detection) and **phase 2** (schema classification). Specifically, it:

- Defines custom Dataset classes: creates PyTorch Dataset subclasses for each phase (Collobert et al., 2002).
- Encodes input texts: uses tokenizer[1] to convert textual data into token IDs, attention masks, and other necessary inputs for the models.
- Processes labels appropriately: prepares and formats labels according to model requirements (e.g., BIO labels for token classification and multi-hot vectors for multi-label classification).
- Prepares data for training: organizes the data into a format compatible with the Hugging Face Trainer API (Wolf et al., 2020).

For the boundary detection task, a training set of 3,834 samples and a validation set of 427 samples are used. On the other hand, the schema classification task relies on a training set of 2,851 samples and a validation set of 317 samples.

At the core of the boundary detection module is a token-level classification model based on BERT (Devlin et al., 2017). This model is designed to assign BIO (Begin, Inside, Outside) labels to each token within a given text sequence, thereby determining the boundaries of relevant chunks. Mathematically, for an input token sequence $T = \{t_1, t_2, \ldots, t_n\}$, the model computes contextual embeddings $E = \{e_1, e_2, \ldots, e_n\}$ using the transformer. These embeddings are then passed through a linear layer W and a softmax activation function to obtain the probabilities for the BIO labels of each token:

$$P(l_i \mid t_i) = \text{softmax}(We_i),$$

where l_i denotes the label assigned to token t_i.

In parallel, the schema classification component is designed as a multi-label classification model, which uses a similar transformer-based architecture. Unlike the boundary detection model, this classifier operates at the sequence level, taking

[1] A tokenizer is a tool which transforms raw text into smaller segments known as tokens – such as words, subcategories of words or characters, and then converts those tokens into numeric identifiers which the model can understand and process.

entire text chunks as input and simultaneously predicting the presence of multiple schema categories. For a given input chunk $C = \{c_1, c_2, \ldots, c_m\}$, the model generates an encoding h_{CLS} based on the [CLS] token, which represents a summarized representation of the entire segment. This representation is then passed through a dense layer followed by a sigmoid activation function, producing probability values that a particular schema is present:

$$P(s_j \mid C) = \sigma(Wh_{CLS} + b),$$

where s_j represents a specific schema and σ is the sigmoid activation function. The boundary detection model measures error using categorical cross-entropy, while the schema classification model uses binary cross-entropy (Devlin et al., 2017). Mathematically, the interaction between the two models can be represented as:

1. Chunk detection: $\{T_1, T_2, \ldots, T_n\} \rightarrow \{C_1, C_2, \ldots, C_k\}$
2. Schema classification: $\{C_1, C_2, \ldots, C_k\} \rightarrow \{S_1, S_2, \ldots, S_k\}$

The models are further fine-tuned using the AdamW optimizer with a learning rate of 5e-5, achieving a balance between convergence speed and stability (Kingma & Ba, 2017). Training was conducted over 50 epochs[2] with a batch size[3] of 32, optimizing computational efficiency without compromising model performance. Gradient accumulation[4] was applied to effectively utilize the batch size within the constraints of available GPU memory, ensuring that the models could process a sufficient amount of data per update step. To prevent overfitting[5] and improve generalization, both models incorporated dropout regularization with a dropout rate of 0.18.

5. Model performance

The model's performance was evaluated using a set of metrics covering precision and recall, providing a balanced assessment of its capabilities. On the validation set, the boundary detection model achieved the following results:

- Evaluation Loss: 1.459
- Precision: 0.309
- Recall: 0.347
- F1 Score: 0.327

[2] An epoch represents the complete processing of the entire data group used for training, which means that during each epoch the model sees all the examples from the training set once.

[3] A batch size of 32 means that 32 examples are processed at the same time before the model parameters are adjusted, which allows for work with large datasets in smaller, more easily manageable segments.

[4] Gradient accumulation adds the gradients from several smaller batches before it applies a single update. Thus the model can simulate a greater effective batch size without the need for additional GPU memory.

[5] Dropout regularization randomly excludes a certain percentage of neurons in a network during training – in this instance 10%, to avoid allowing the model to become overdependent on any individual group of links. This improves the overall robustness of the model.

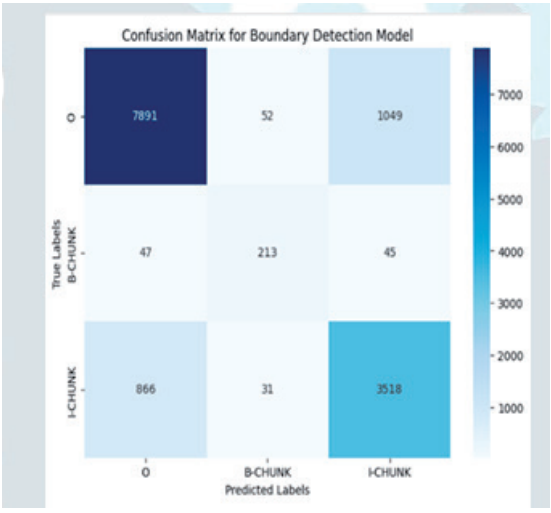
These values indicate a moderate level of performance, with room for improvement in accurately detecting chunk boundaries. The relatively low precision and recall suggest challenges in minimizing false positives and false negatives. The confusion matrix is provided in Appendix 1.

In contrast, the schema classification model demonstrated impressive performance on the validation set, achieving the following results:

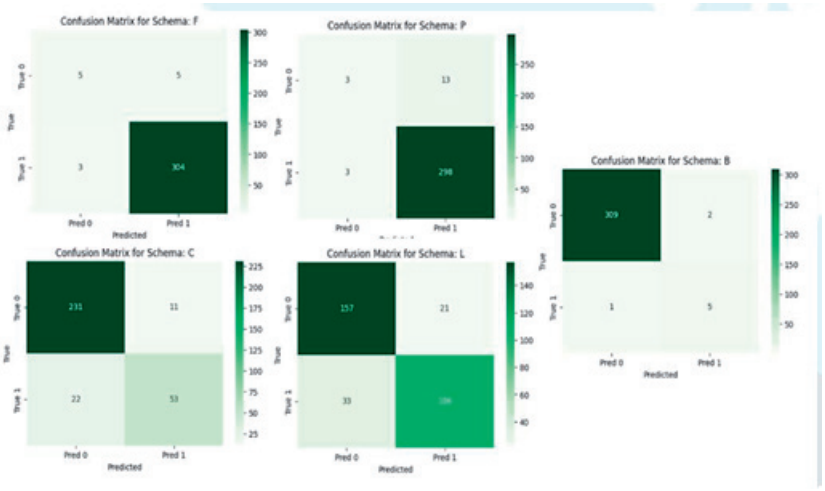
- Evaluation Loss: 0.356
- Precision: 0.936
- Recall: 0.925
- F1 Score: 0.930
- Accuracy: 69.71%

The high levels of precision, recall, and F1 score reflect the model's robust ability to accurately assign multiple relevant schema labels to each text chunk. These metrics indicate the model's expertise in handling multi-label classification tasks, effectively balancing the trade-offs between precision and recall to achieve strong overall performance. Confusion matrices for all five schema categories are provided in Appendix 2.

6. The discussion and further studies



Appendix 1: Confusion Matrix for the Boundary Detection Model



Appendix 2: Confusion Matrices for the Schema Identification Model

These two models are combined as follows: for any new input sequence of tokens, regardless of its length, the segmentation model first extracts meaningful chunks that carry schemas. For example, for the hypothetical input sentence chain:

"He fell into a hole. He moved away from it. Then, he went into a house." the model returns the following annotation:

- **Chunk 1:** he fell into a hole
  Identified schemas: P, F, C
- **Chunk 2:** he moved away from it
  Identified schemas: P, F, L
- **Chunk 3:** then, he went into a house
  Identified schemas: P, F, C

After optimizing the boundary detection and schema classification models, future directions of this research are focused on developing advanced models that integrate scaling modifiers and specifiers, thereby enriching the semantic depth and contextual precision of the entire system. The scaling model is designed to complement the identified schemas with quantitative adjustments, such as positive or negative indicators (+/-), which denote the magnitude and direction of the schema's valence. By integrating these modifiers, the scaling model aims to increase the granularity of schema annotations and to enable subtler interpretations and applications that require precise quantitative data.

At the same time, a specification model has been designed to refine and contextualize existing schemas. This model aims to add modifiers such as "up," "down," "end path," and similar qualifiers, providing schemas with additional layers of semantic information, more detailed and context-aware classification, and flexibility for tasks requiring a high degree of specificity and adaptation.

However, the current architecture shows significant limitations, particularly in the boundary detection component, where the Token-Level F1 Score was 0.3067.

This result indicates moderate accuracy in chunk boundary identification, while the low precision and recall suggest a tendency toward false positives and false negatives. Since this model passes the formed chunks to subsequent processing, these errors can propagate to the next models, including the scaling and specification models. Addressing these challenges requires:

- Extended training regimes: Longer training periods, iterative hyperparameter optimization, and advanced regularization techniques to improve accuracy and generalization capabilities of the models.
- Increasing the dataset: Incorporating more diverse and representative samples in the training set to mitigate overfitting and enhance the model's ability to generalize across different textual contexts.
- Advanced techniques: Using cross-validation and ensemble learning to further strengthen the robustness of the segmentation model.

Furthermore, integrating the scaling and specification models into the existing pipeline requires a cohesive architectural framework that ensures smooth data flow and interoperability between components. This entails developing sophisticated data preprocessing procedures capable of handling additional layers of annotations, such as applying the SCALE schema to other schemas. From an architectural perspective, employing modular and scalable design principles will be crucial to accommodate the growing complexity and interdependencies of the expanded pipeline structure.

References

Antović, M., Jovanović, V. Ž., & Figar, V. (2023). Dynamic schematic complexes: Image schema interaction in music and language cognition reveals a potential for computational affect detection. *Pragmatics & Cognition, 30*(2), 258–295.

Bennett, B., & Cialone, C. (2014). Corpus Guided Sense Cluster Analysis: a methodology for ontology development (with examples from the spatial domain). *Formal Ontology in Information Systems*, *267*, 213–226. https://doi.org/10.3233/978-1-61499-438-1-213

Bird, S. (2006, July). NLTK: the natural language toolkit. In Curran, J. (Ed.) *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions* (pp. 69–72). https://doi.org/10.3115/1225403.1225421

Brownlee, J. (2020, August 17). *Ordinal and one-hot encodings for Categorical Data*. MachineLearningMastery.com. https://machinelearningmastery.com/one-hot-encoding-for categorical-data/ Chardet. PyPI. (n.d.). https://pypi.org/project/chardet/

Clausner, T. C., & Croft, W. (1999). Domains and image schemas. *Cognitive Linguistics, 10*(1), 1–31. https://doi.org/10.1515/cogl.1999.001

Collins, G., & Syme, D. (1995). A theory of finite maps. In E. T. Schubert, P. J. Windley, & J. Alves-Floss (Eds.), *Higher Order Logic Theorem Proving and Its Applications: 8th International Workshop Aspen Grove*, *UT*, *USA*, *September 11–14*, *1995 Proceedings* (pp. 122–137). Berlin: Springer.

Collobert, R., Bengio, S., & Mariethoz, J. (2002). *Torch: a modular machine learning software library*. IDIAP RR 02-46. Lugano: Dalle Molle Institute for Perceptual Intelligence.

Dahlgren, K. (1988). *Naive semantics for natural language understanding*. Norwell/ Dordrecht: Kluwer Academic Publishers.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, May 24). *Bert: Pre training of deep bidirectional Transformers for language understanding*. arXiv.org. https:// arxiv.org/abs/1810.04805

Dodge, E., & Lakoff, G. (2005). Image schemas: From linguistic analysis to neural grounding. In B. Hampe (Ed.), *From Perception to Meaning* (pp. 57–92). Berlin: Mouton De Gruyter. https://doi.org/10.1515/9783110197532.1.57

Evans, V., & Green, M. (2006). *Cognitive linguistics: An introduction*. Edinburgh: Edinburgh University Press.

Gibbs Jr, R. W., Beitel, D. A., Harrington, M., & Sanders, P. E. (1994). Taking a stand on the meanings of stand: Bodily experience as motivation for polysemy. *Journal of Semantics, 11*(4), 231–251.

Gibbs, R. W., & Colston, H. L. (1995). The cognitive psychological reality of image schemas and their transformations. *Cognitive Linguistics, 6*(4), 347–378. https://doi. org/10.1515/cogl.1995.6.4.347

Gromann, D., & Hedblom, M. M. (2017). Kinesthetic mind reader: A method to identify image schemas in natural language. In P. Langley (Ed.), *Advances in Cognitive Systems, 5*, Paper 9. Cognitive Systems Foundation. https://dagmargromann.com/ files/ACS_final_2017.pdf.

Helbig, H. (2014). *Knowledge representation and the semantics of natural language*. Berlin: Springer.

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

Kapetanios, E., Tatar, D., & Sacarea, C. (2013). *Natural language processing: semantic aspects*. Boca Raton: CRC Press.

Kingma, D. P., & Ba, J. (2017, January 30). *Adam: A method for stochastic optimization*. arXiv.org. https://arxiv.org/abs/1412.6980

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

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York City: Basic Books.

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.

Mandler, J. M. (2004). *The foundations of mind: Origins of conceptual thought*. Oxford: Oxford University Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of machine Learning research, 12*, 2825–2830.

Van Rossum, G. (2020). *The Python Library Reference*, release 3.8.2. Python Software Foundation.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017, June 30). *Attention is all you need*. arXiv.org. https://arxiv.org/abs/1706.03762v4

Wachowiak, L. (2020). Semi-automatic Extraction of Image Schemas from Natural Language. In L. Gschwandtner et al. (Eds.), *Proceedings of the MEi:CogSci Conference 2020* (p. 105). Bratislava: Comenius University.

Wachowiak, L., & Gromann, D. (2022). Systematic analysis of image schemas in natural language through explainable multilingual neural language processing. In N. Calzorali et al. (Eds.), *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 5571–5581). International Committee on Computational Linguistics.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., … Rush, A. M. (2020, July 14). Huggingface's transformers: State-of-the-art natural language processing. Sydney: Association for Computational Linguistics. Retrieved from /https://arxiv.org/abs/1910.03771

THE SASE JOURNAL
ISSN 3042-2930, UDK 821.111

№ 2, 2026
pp. 53-77

Original research paper
UDC 781.1:81'342.1
DOI: https://doi.org/10.46630/sase.2.2026.1
ORCID: 0009-0000-7999-3786

THE SPEAKING BOW: LINGUISTIC RESONANCES IN STRING PLAYING

Denise Fan[1], Postdoctoral Research Fellow,
Academy of Music – Hong Kong Baptist University, Hong Kong

Abstract

The phenomenon of subconscious linguistic-prosodic imprint has been studied in music composition (Patel & Daniele, 2003; Temperley, 2022), yet its operation in performance remains terra incognita. Such a persistent gap reflects the formidable challenge of assessing how native phonetic patterns covertly shape the "feel" in musical expression – idiosyncratically encoded in articulation. To investigate this embodied interaction, the present study pioneers a novel methodology that deploys the unparalleled articulative freedom offered by the Baroque bow to distill language "flavors." Multilingual adaptations of Toshinobu Kubota's *La La La Love Song* serve as phonetic templates for translating distinctive prosodic features, including Japanese pitch accent, Cantonese syllable-finals, and English stress-timing, into repeatable bowing schemata. Preliminary spectrogram analyses suggest that string players' articulation aligns with native prosody, particularly in speech-inflected performance. For instance, spectral qualities in Anner Bylsma's strokes mirror Dutch guttural timbres, and Ophélie Gaillard's note shaping replicates the intensification of French final lexical stress. The proposed schemata provide a blueprint for interrogating "native-ness" and "foreign-ness" in performance. By mapping phonetic patterns onto bowed gestures, this work offers a method to analyze the permeation of native prosody in instinctive playing, refining expressive intent through the conscious articulation of the sensuality inherent in linguistic sound.

Keywords: speaking bow, musical articulation, linguistic acoustic patterns, language-music interplay cultural resonance

[1] Email address: denisecyfan@gmail.com
Corresponding address: Flat H, 10/F, Block 16, Chi Fu Fa Yuen, Pokfulam, Hong Kong

1. Introduction

Starker definitely has a tendency to perceive phrases from the perspective of his native language. That suits how he plays, how he thinks about music, and how it fits into his overall concept of how music is put together. This is organic (Geeting, 2008: 99). Geeting's penetrating perspective chimes with a persistent intuition among performers: that native language silently shapes the musical mind. By "organic," she calls attention to the phenomenon of how deeply music and language are integrated as an "expressive whole" to facilitate communication (Faudree, 2012: 520). For Starker, this blend of auditory experience "in the head" is projected outward through cello playing. Wordless yet profoundly linguistic, such expression powerfully articulates the rich complexity of human interiority. The mystery of this alchemy has been a time-honored philosophical fascination, driving scholars to map the structural conventions of music onto those of language, fueling countless arguments for their isomorphic parallels or categorical distinction (Temperley, 2022).

The plausible processes involved were teased out by Bernstein (1973/1976) in his provocative *Norton Lectures, The Unanswered Question*, where he compared musical units with three generative topics in linguistics, namely phonology, syntax, and semantics. Deploying deeply premeditated analogies and metaphors, his concepts require a highly erudite and hermeneutical mind to apprehend the unconscious logic behind music composition as demonstrated by human's faculty in formulating words and sentences. In an article that contravenes Bernstein's theory, Keiler (1978: 195) acknowledges "the intrinsic interest and value" of the subject. I argue that there is an obvious reason for this, as inscribed in Geeting's remark with the word "native." There is a particular "ring" to this word, especially when contrasted with its shadow term, "foreign." This visceral tension felt in music was noted by Robert Hall (1953/1972: 284), "the Englishman simply feels an 'instinctive' affinity to Elgar's music, and the non-Englishman feels its 'strangeness'." He ascribed the subconscious preference to the flavor of British English, characterized by "a wide range of variation in pitch and a predominance of falling patterns," which permeated Elgar's works to the extent that even the signature minor-third drop in the Welsh tune of his *Introduction* and *Allegro for Strings*, Op. 47, was irresistibly developed through his trademark leaps and descents.

The musical "native" and "foreign" were most sharply exposed by the serious discourse on Hungarian rhythm (Hooker, 2013: 154–229), epitomized in Bartók's and Kodály's compositions born out of their exhaustive folksong research. Their targeted effort to embed native speech patterns into musical structures transcends mere nationalism, revealing something more fundamental: the linguistic identity etched into every thought, for indeed "every human being speaks all the time to himself when he is 'thinking silently'" (Hall, 1953/1972: 284).

This inextricable infiltration of linguistic prosody into music was first empirically validated by Patel and Daniele (2003). Applying the nPVI (normalized Pairwise Variability Index) to instrumental themes by native English- and French-speaking

composers from the turn of the twentieth century – an era when the search for national expression illuminated linguistic-musical connections, they demonstrated how the musical rhythms mirrored the characteristic stress- and syllable-timed patterns of the two languages, respectively. Thereafter, studies proliferated in examining the cognitive intertwinement of speech and musical prosody (pitch, rhythm, and stress) as a means to trace the culturally constructed sonic "mental framework" (Patel, 2008: 9) that shapes our auditory predispositions (Arbib, 2013; Patel, 2008; Scharinger & Wiese, 2022).

While approaching language and music as "two distinct sound systems" (Patel, 2008: 9) for comparison has yielded fruitful insights into how their interaction forges an acoustic signature that calibrates our auditory perception, its manifestations are best studied not in the musical score, but in the embodied utterance through performance, which renders our "accent" audible. Defined as "a way of pronouncing a language that is distinctive to a country, area, social class, or individual" (Oxford University Press, n.d., Definition II.7.a), an accent is a cultural construct where any minor variations are instantly detected to place a speaker as "'one of ours' or 'not one of ours'" (Graham, 1969: 448). Since this installed (though continuously evolving) auditory filter colors our hearing constantly, it inevitably transfuses instrumental playing via action-perception coupling (see Novembre & Keller, 2014). In this light, the observed phenomenon that "a musician tends to reproduce performances of the same music with the same prosodic choices" (Palmer & Hutchins, 2006: 247–248) exemplifies this very transfusion.

One might argue that a performer's accent, discernible in their consistent musical prosody, is developed through training and sharpened by in-depth, diverse stylistic acquisitions. Following this conventional view, "native-ness" or "foreign-ness" is gauged by whether the performer possesses sufficient knowledge to express the spirit of an artistic syntax, a point compellingly made in cellist Guy Fishman's (2014) critique of technically accomplished students for "speaking a language with a heavy accent." Yet, before an accent is a matter of stylistic fluency, it is first and foremost a product of the cognitive ear. Its "native-ness" originates in the infant's processing of the "music-ness" of speech (Brandt, Gebrian, & Slevc, 2012). By analyzing how linguistic prosody seeps into instrumental performance, we can pinpoint the specific nuances in playing informed by a performer's native language.

Tracking this influence has remained a profound challenge, with virtually no existing literature to build upon. Precisely because this missing piece reveals the "tendency" (Geeting, 2008: 99) that shapes our aesthetic identity, it demands a lens exquisitely honed in on the nuances of articulation, one we find in the practice of Baroque string playing.

2. Sound for words to Sound of words

The essence of Baroque music lies in evoking the affective dimensions of any given text. The power of words is crystalized in Leibniz's philosophy that "language

is not the vehicle of thought but its determining medium" (Steiner, 1975: 74). In pursuit of "the incorruptible truth" (Egginton, 2009: 10), momentarily accessed through the state of awe, Baroque composers attempted to evoke the soul's yearning for the truth by tuning into the sense and sensuality of a text. By acknowledging the capability of words to carry the depth of human felt passions − through the synthesis of intellect and imagination − composers made the intelligibility of the text their top priority, giving rise to the *seconda pratica*. The shift to plangent monody, harsh chromatic punctuations, and storytelling through harmonies all served to create a resonance chamber for the words to "penetrate the intellect of others" (Caccini, 1602/2021: 27).

Caccini enacted this priority by radically reinterpreting Plato's view that a song consists of "the words, the tune, and the rhythm," and "the music and the rhythm must follow the speech" into the dictum, "music was nothing other than the words, and [then] the rhythm, and finally the sound, and not otherwise" (Caccini, 1602/2021: 26–27). In doing so, he rendered music a mere sonic vessel for words. This reorientation epitomizes the Baroque desire to stir the heart with passions that remain, for the mind, "just beyond grasp" (Egginton, 2009: 11) − even though the music thereby created is yet another "corrupted" (Egginton, 2009: 10) form of mediation, one that deceives by dramatizing the illusion of promising a truth.

The Baroque quest for a universal musical language that rivals persuasive speech to transport audience to affective states springs a multitude of creative experimentations, which enthralled generations of "Neobaroque" (Egginton, 2009: 16) pursuers. United by "a love for language" (Fishman, 2014), contemporary HIP (historically informed performance) practitioners approach this musical speech act via three interconnected fronts: literacy, inflection, and timbre. A sophisticated literacy has been honed by an incessant research into stylistic markers at the regional level (e.g. Roman vs. Venetian; see Vanscheeuwijck, 2020), providing practical insights into interpreting individual composers' works. The prevailing stratified inquisitions into each composer's unique language were recast as the bedrock principle for learning historical performance, which is "just another language" (The Juilliard School, 2016). This "foreign language" (Fishman, 2014) is distinguished by a highly inflected delivery that accords with the strategic combinatorial discourse (see Quantz, 1752/1966). Its mastery is fundamentally what rhetoric is all about, which Rifkin argues is simply "a way of describing what good musicians of any era have always done" (Sherman, 1997: 172). To internalize the prescribed inflections, players turn to period instruments to explore the specific articulations they afford, impressed by their palpable timbres. The physical engagement with these instruments and the consequential "auditory mental imagery" (Barbero & Calzavarini, 2024) conceived through the "grain" of the sounding body (Hui, 2020: 3) vividly coincide with speaking in the mind. Hence, the obsession with the word "speak" − a near-ubiquitous term within HIP discourse.

Carrying on the humanistic adventure sparked by the Baroque "theater of truth" (Egginton, 2009), a logical Neobaroque exploration would be to confront the very sound of words as we navigate the digital matrix of babel in our everyday lives.

This mass of sounds has severely numbed our precious sensitivity to subtle nuances in communication, transmitted through years of enculturation as linguistic prosody. Unveiling prosodic wiring gives us a roadmap to discover characteristic pathways to move the "affections of the soul" (Haynes & Burgess, 2016: 130). I propose we leverage the fruits of HIP practice and begin the conversation by first tracing the sonic aura of "different tongues," achievable by the speaking Baroque bow.

3. Bowing schemata

Consider the musical snippet shown in Figure 1 – once it sounds in the head, it most likely prompts an immediate "aha" replay. This could manifest as a hum with every note unquantifiably shortened and animated, or as a carolling of *For unto us a child is born* from Handel's *Messiah*. Regardless of which mental replay occurs, the final eighth-note pair, C and B, is decisively treated as a cohesive gestural unit: a slight accent is applied to the C, followed by a decrescendo through to the B, which is released.

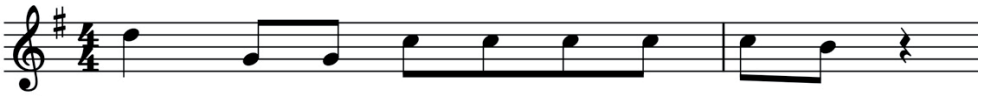


Figure 1: Musical theme of *For unto us a child is born*

This expressive microcosm (stress-connect-diminish-release) feels natural since both the prosodic and semantic aspects of the word "born" are satisfactorily fulfilled by the accented appoggiatura C resolving to B (Figure 2).

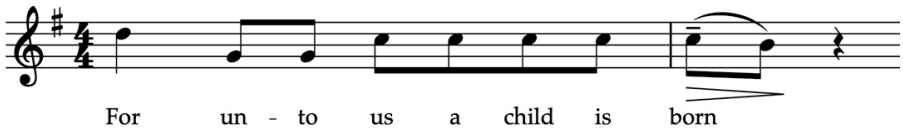


Figure 2: *For unto us a child is born* (theme with expressive markings)

The other notes of the theme are articulated in a varied, unquantifiable manner, owing primarily to two factors: first, the strongly emphasized "for" (by means of its metrical placement on a strong beat and its setting as a declarative high note marked off by a perfect-fifth leap) is an unstressed word in normal speech, thereby inviting ambiguous lengths and intensities; second, some may treat the entire phrase as a single unit with a primary stress on "born," while others may segment it into "for unto us, a child is born." This reading is musically supported by the leap from G to C on "-to us," a gesture that encourages a "scoop" into the first C ("us") – somewhat shortening it – and that naturally invites a secondary emphasis on the third C ("child").

The unanimous treatment of "born" and the multiple interpretive possibilities offered by what comes before demonstrate both the robust prosodic imprint as well as the fluctuating potentials imparted by the linguistic system, especially when the semantic meaning and the prosody of the text are congruent or incongruent with the musical design. In fact, the music of the chorus *For unto us a child is born* was adapted from Handel's earlier duetto, *No, di voi non vo'fidarmi* ("No, I do not want to trust you"; Figure 3). This original version presents a clear phrasal directive: an emphatic, resentful "no" followed by a single teleological gesture with an arrival on the syllable "dar."

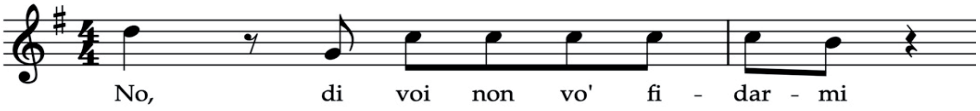


Figure 3: Musical theme of *No, di voi non vo'fidarmi*

The precise linguistic-musical prosodic alignment of the Italian original (*No, di voi non vo'fidarmi*) and the tension and acceptance revealed by the English adaptation serve as a compelling validation of Mattheson's idea that the notes themselves hold no intrinsic expressive value but are affectively charged by words (Haynes & Burgess, 2016: 188). More crucially, this commutable melodic pattern illuminates how the distinctive acoustic properties are highlighted in subsequent versions in other languages, providing a unique window for performers to explore the characteristic "flavors" of different languages.

3.1. Experimental foundation and methodology

To this end, I selected Toshinobu Kubota's (1996) *La La La Love Song* (best known as the theme song for the drama series "Long Vacation") as an experimental template for two reasons. First, its R&B tune counterbalances classical stylistic predispositions. Second, the song exists in multiple adapted versions spanning a diverse array of languages. From this corpus, I undertook a close analysis of the original Japanese rendition, alongside versions in my native Cantonese, as well as in English, Mandarin Chinese, and Malay.

The experimental process relied essentially on the specific affordances of the Baroque bow and cello strung with gut. The bow's convex shape and tapered design produce a pronounced decrescendo in weight from frog to tip, making it feel like a natural extension of the arm and enabling the player to "speak" with unparalleled articulative freedom. Precise tactile engagement is further demanded by the gut strings, which are acutely sensitive to bow pressure, speed, and contact point. Coupling both allows a high degree of precision in articulation, effectively mirrors verbal utterances by marking, shaping, and closing musical units from individual note to phrase level.

While the articulative freedom offered by the Baroque bow encourages obsessive sound copying down to the micro-phoneme level, as illustrated in Reiter's guide for translating Italian pronunciation into violin bowing:

> In "Corelli," the three syllables are not equal either in length or stress: "Co" is accented with an immediate diminuendo leading into the rolled "r." There is a Messa di voce (<>) through the "e," the climax of which is the strongest part of the word, and there is a little extra bounce on the "lli." An Italian speaker would pronounce both "l"s, with a barely perceptible break in the sound between the "e" and the "lli." (Reiter, 2020: 32)

The overly meticulous approach is not only impractical but also incapable of reflecting certain phonetic differences such as "bet" vs. "bat" and "bet" vs. "debt." Therefore, I developed bowing schemata designed to capture one quintessential prosodic feature for each target language.

Verse lines were carefully chosen both to ensure they were set to identical melodic material across all language versions and to avoid non-lexical vocables (e.g., la la la, na na na). Furthermore, melodic and rhythmic normalization was performed to remove minor discrepancies that would detract attention from articulation. The non-linear experimental process involves several key steps:

- Repeated recitation at various speeds to discern phonetic features that either "roll off the tongue" or cause articulatory resistance, as well as to internalize prosodic nuances
- Utilizing Speechify, a text-to-speech AI tool, to learn the pronunciation of words in unfamiliar languages (particularly Malay) and to observe the varied inflections generated by different speaker styles
- Cross-verifying my internal auditory representation of pronunciation with various prosody checkers, dictionaries, and IPA (International Phonetic Alphabet) resources
- Testing and distilling prosodic features on the cello, using audio recording for evaluation to determine which elements can be reliably captured and recognized
- Annotating the lyrics and transferring corresponding expressive realizations onto the music score to encode prosodic intent; devising shorthand notation as necessary to facilitate performance.

Imperatively, I relied on my intuition and practical experience to evaluate the feasibility of executing the identified phonetic features on the cello. The resulting bowing schemata are inevitably a form of abstract representation of the embodied acoustic apperception.

3.2. Japanese

- Distinctive phonetic feature: Pitch accent
- Articulation was achieved in three steps:

1. Grouping morae: Morae were grouped into syllabic units and slurred under a single bow stroke to reflect their cohesive articulation.

ためいきのまえに
ここにおいでよ
いきがとまるくらいの
あまいくちづけをしようよ
ひとこともいらないさ
とびきりのいまを

Figure 4: Japanese verse lines with moraic groupings underlined

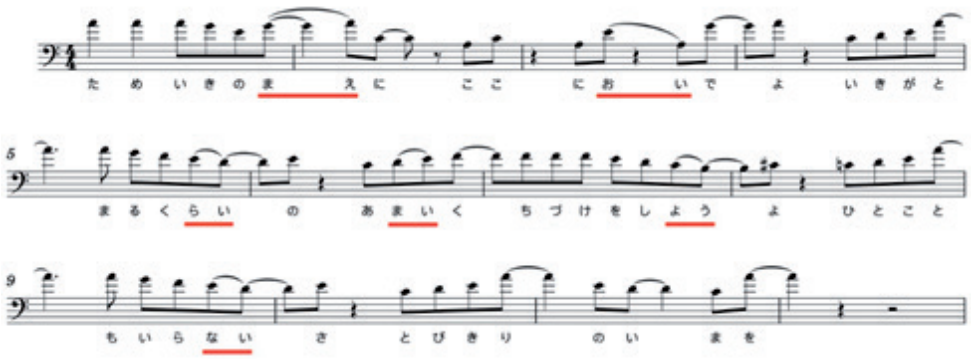


Figure 5: Score excerpt illustrating moraic groupings realized with slurs

2. Mapping pitch accents: Each pitch-accent type was mapped to a bowing sequence that corresponds to the pitch contour.

Pitch-Accent Type[4]	Pitch Contour	Bowing Sequence
heiban (“flat board”; “accentless”)	low-flat	up > down (even pressure)
atamadaka (“head high”)	high-low	down > up
nakadaka (“middle high”)	low-high-low	up > down > up
odaka (“tail high”)	low-high	up > down

Figure 6: Corresponding bowing sequences for the four Japanese pitch-accent types

ためいきのまえに
ここにおいでよ
いきがとまるくらいの
あまいくちづけをしようよ
ひとこともいらないさ
とびきりのいまを

Figure 7: Japanese verse lines annotated with pitch accents

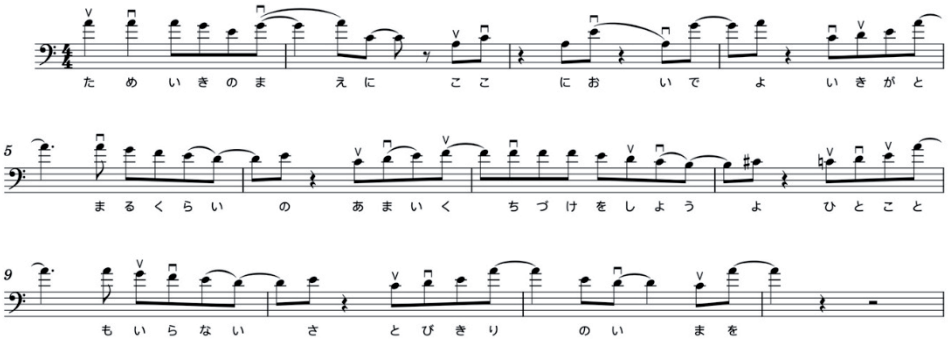


Figure 8: Score excerpt with bowing indicated in accordance with pitch accents

3. Phrasing: Word-level pitch accents were integrated into sentence-level prosodic contours, shaped by phrasal dynamics.

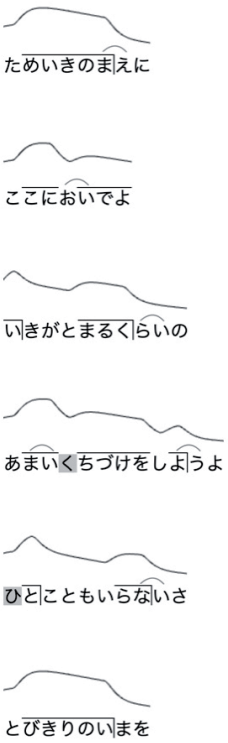


Figure 9: Japanese verse lines with pitch contours machine-generated by Prosody Tutor Suzuki-kun

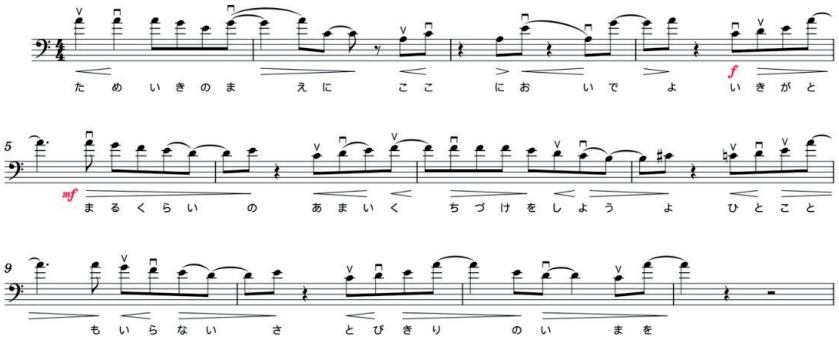


Figure 10: Score excerpt with dynamic markings mapping pitch-accent contours
**Refer to Audio 1 for performance realization.

3.3. Cantonese

- Distinctive phonetic feature: Syllable-final types
- Shorthand symbols were devised to facilitate intuitive bow articulation in real time for each syllable-final type. The corresponding bow stroke qualities are detailed in Figure 11.

Syllable-Final Types	IPA Finals	Shorthand Symbols	Bow Stroke Qualities
monophthong	:	—	flat
diphthong	:i̯ :u̯	⌒	swell (with a down-bow)
nasal coda	:m :n :ŋ	⌟	nudge (slight pressure with a slow bow)
plosive coda	:p :t :k	ʼ	short and sharply articulated (with an up-bow)

Figure 11: Cantonese syllable-final types with devised shorthand symbols and corresponding bow stroke qualities

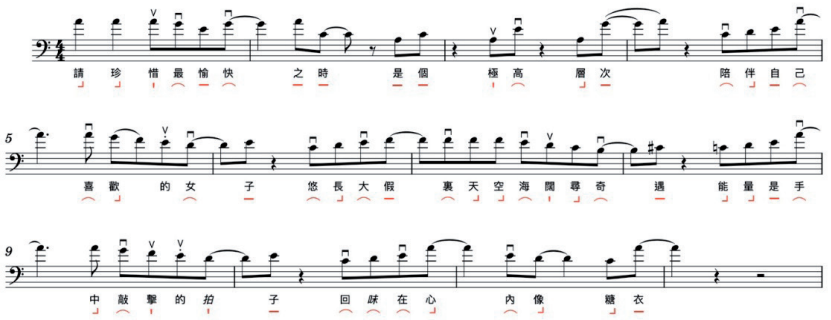


Figure 12: Score excerpt annotated with shorthand symbols and corresponding bowing
*The particle “的” was treated with a light up-bow staccato, reflecting its reduced prosodic weight.

*In keeping with conventions of bowing economy, the devised sequences were occasionally reversed (words italicized) to ensure practical playability.
**Refer to Audio 2 for performance realization.

3.4. English

- Distinctive phonetic feature: Stress-timed prosody
- Two levels of prosodic stress were captured: lexical (word-level) and phrasal (sentence-level). While lexical stress is relatively fixed, phrasal stress requires semantic interpretation, which varies by performer. In both cases, stressed elements were articulated with a down-bow. Brackets were used to visually group notes belonging to a single word.

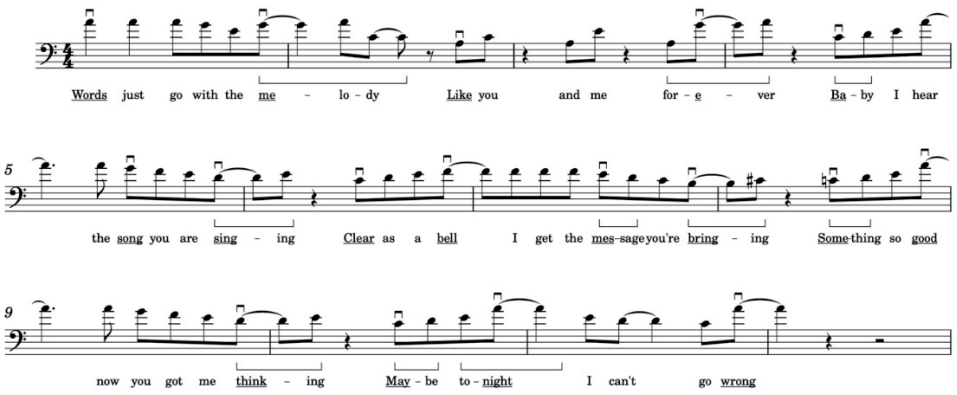


Figure 13: Score excerpt showing down-bow stresses and word groupings
**Refer to Audio 3 for performance realization.

3.5. Mandarin Chinese

- Distinctive phonetic feature: Tone
- Shorthand symbols were devised to map the four tones to their corresponding bow stroke qualities.

Tones	Shorthand Symbols	Bow Stroke Qualities
1 : –	—	even
2 : /	⌣	scoop
3 : v	▾	dense press
4 : \	⩾	accented with a sustained decay

Figure 14: Bow stroke qualities matching the four tones in Mandarin Chinese, represented by shorthand symbols

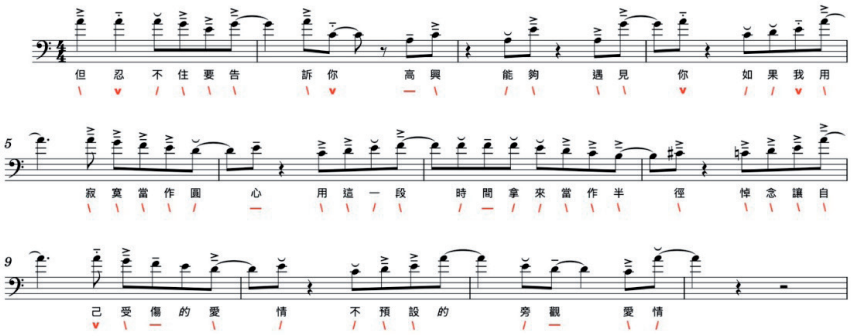


Figure 15: Score excerpt showing shorthand articulation markings in conjunction with the four tones in Mandarin Chinese

*Tonal variations can exist depending on the speaker.

*The particle "的" (italicized) was played with a light bow stroke.

**Refer to Audio 4 for performance realization.

3.6. Malay

- Distinctive phonetic feature: Syllable-onset articulation
- Shorthand symbols were devised to express the bow stroke qualities in imitation of the four major types of syllable-onset articulation in Malay. Brackets were used to visually group notes belonging to a single word.

Types of Syllable-Onset Articulation	Shorthand Symbols	Bow Stroke Qualities
Alveolar [n, t, d, s, l]	⌞	even with a swift start
Labial [m, b, p]	v⁻	slight press (stopped) at the start of stroke
Palatal [ɲ, tʃ, dʒ, j]	⌣	scoop
Velar [ŋ, k, g, w]	▬	marked, dense, tight, with a concentrated core

Figure 16: Bow stroke qualities and shorthand symbols matched to the four major types of syllable-onset articulation in Malay

*Light syllables (e.g., "-an," "a-") falling outside the core articulation types were realized with a simple up-bow stroke to reflect their reduced prosodic weight.

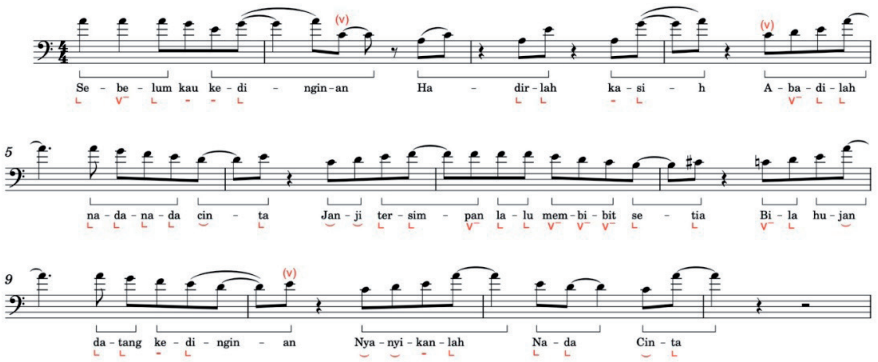


Figure 17: Score excerpt marked with shorthand symbols representing the four major types of syllable-onset articulation in Malay
*Notes under a cohesive syllabic gesture are slurred under one bow.
*Syllables with initial [h] were simply bowed according to musical context.
**Refer to Audio 5 for performance realization.

4. *Non l'intendite parlare?*

Non l'intendite parlare? ("Do you not hear it speak?") is a well-circulated quote within the HIP circle. It is apocryphally attributed to Corelli (see Taruskin, 2023: 379) with the intention of imploring string players to adopt a playing style akin to an eloquent speech − with clarity, nuanced inflections, and a pronounced sense of shape that enlivens every note, musical gesture, and phrase. Among instrumentalists, a perennial debate revolves around whether the instrument should "speak" or "sing," which is ultimately a negotiation between clear diction and resonant tone production. Any first-class musician, of course, marries both much in the manner of the finest singers. That being said, there is an added "unspeakable" layer when HIP musicians insist on a speaking quality, as invoked from the way they hyper-articulate the word "speak." The acoustic charge of the word, loaded with its semantic resonance, induces a phonetic chill that grips the mind.

To visualize this effect, we can compare spectrograms of the opening scale in J. S. Bach's *Prelude* from the C-major Cello Suite as "spoken" by Anner Bylsma (Figure 18; Bach, 1992) and "sung" by Alisa Weilerstein (Figure 19; Bach, 2020). The dynamic blotches of formants in Bylsma's graph stood in stark contrast to Weilerstein's canvas of evenly sustained partials.

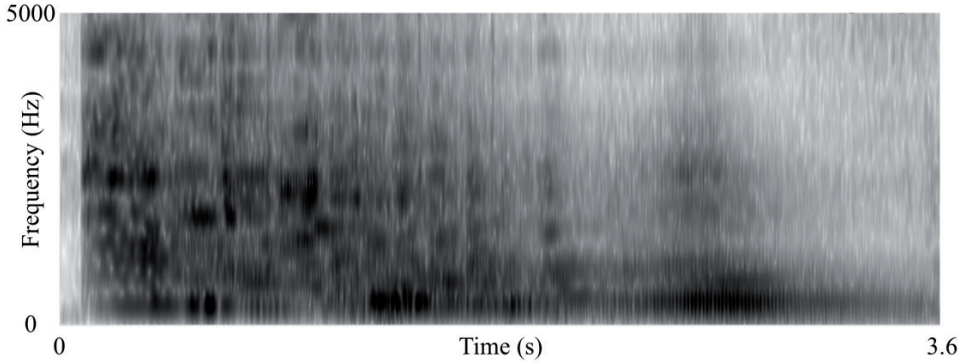


Figure 18: The "spoken" articulation: Spectrogram of the opening scale in J. S. Bach's *C-major Cello Suite Prelude* (Anner Bylsma)

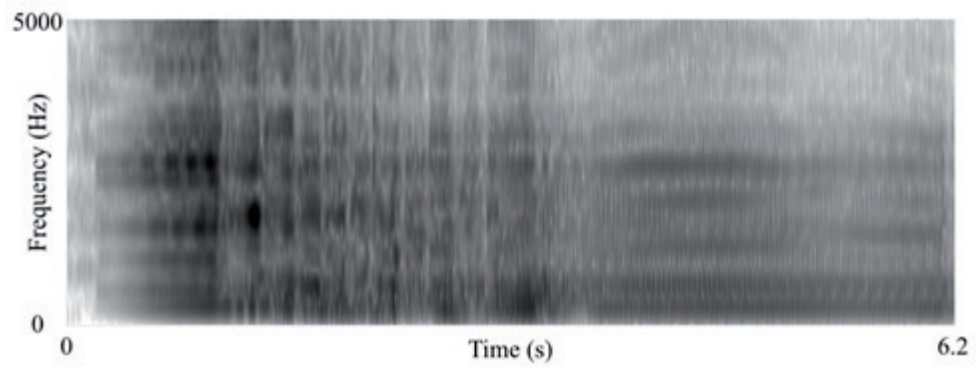


Figure 19: The "sung" line: Spectrogram of the opening scale in J. S. Bach's *C-major Cello Suite Prelude* (Alisa Weilerstein)

Zooming closer into Bylsma's first note C (Figure 20) reveals an intriguing feature: a visible splice in the upper overtone partials. This momentary discontinuity is reminiscent of the voiceless plosive in the guttural Dutch "g" in "goeiemorgen" (Figure 21; Easy Dutch, 2023). Notably, since both the tonic C and the dominant G exhibit this characteristic "stop gap" (Figure 22), it could very well be indicative of an iconic timbre present in Bylsma's inner ear, cultivated from his native language.

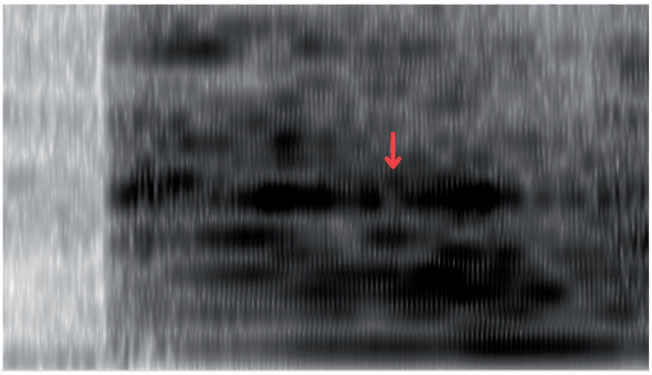


Figure 20: Spectrogram of the articulatory “splice” in Bylsma’s first note C

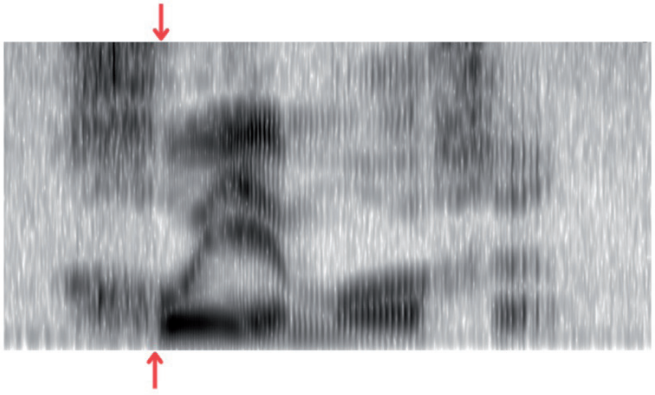


Figure 21: Spectrogram of the guttural “g” from the onset of the Dutch word “goeiemorgen”

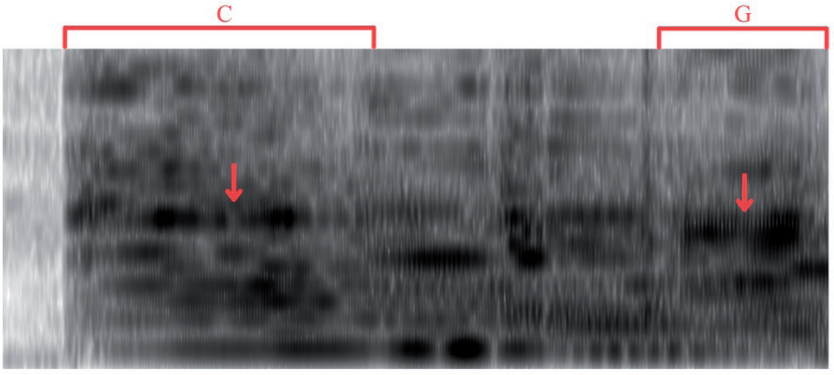


Figure 22: The recurring “stop gap” pattern in both the tonic C and dominant G

Compounding the observed articulatory parallel between bowed expression and voiced utterance, a further exhibit provides compelling support for native prosody steeping into cello playing. Figure 23 presents the spectrogram of the opening C

played by the French cellist Ophélie Gaillard (Bach, 2011), who champions "speaking with the bow" (The Strad, 2022). The gathering of energy intensity towards the end of the note closely resembles the well-established prosodic feature of French final lexical stress (Temperley & Temperley, 2013), as demonstrated by the parallel acoustic behavior in the word "café" (Figure 24; Easy French, 2023).

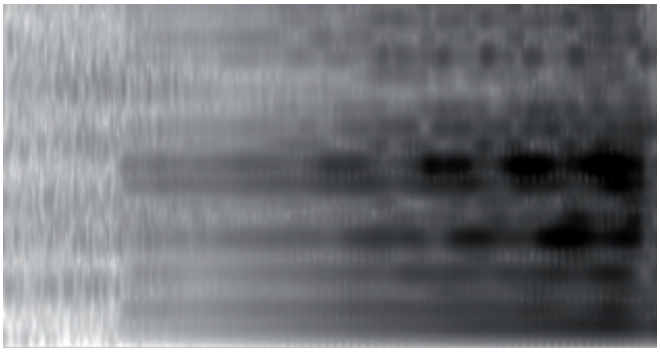


Figure 23: Spectrogram of Gaillard's opening C, showing late-onset energy concentration

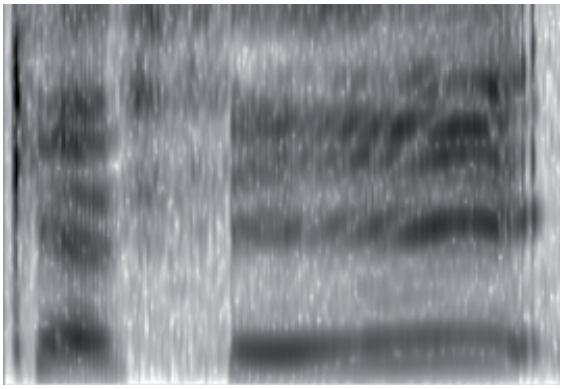


Figure 24: Final lexical stress in the word "café"

Extending the comparative spectrogram analysis to the proposed bowing schemata, it is found that the acoustic profiles of the translated syllable finals in the Cantonese recorded specimen best match the spectrographic profiles of the corresponding syllable-final phonemes found in natural speech (Easy Languages, 2021). The open-ended monophthong displays stable, well-defined formant bands, reflecting its sonorous quality (Figures 25 & 26). Without dramatic pitch changes, the bowed diphthong mirrors the C-shaped formant glide of its spoken counterpart through a gradual increase in intensity that steers the formants upward, effectively giving the impression of the vowel transition (Figures 27 & 28). Figure 29 shows the successful replication of the intended nudge quality for the nasal coda via damped resonance. The energy distribution and the clear demarcation mirror the spoken version (Figure 30) with striking congruence. Although the muted part of the spoken

plosive coda (Figure 32) is not realized in playing for musical reasons, the quick clearance of upper harmonics − reinforced by a loud, short ring in the higher partials − effectively simulates its crisp articulation (Figure 31).

Figure 25: Spectrogram of the bowed monophthong word “子”

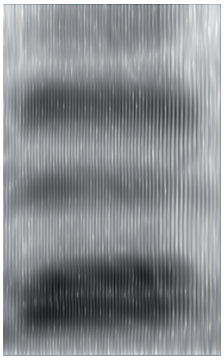


Figure 26: Spectrogram of the spoken monophthong word “喝”

Figure 27: Spectrogram of the bowed diphthong word “陪”

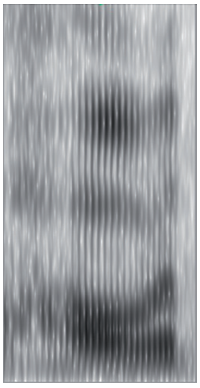


Figure 28: Spectrogram of the spoken diphthong word “好”

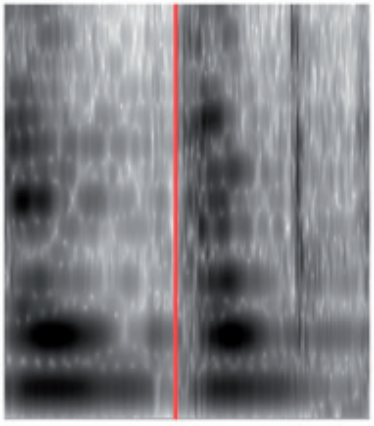


Figure 29: Spectrogram of consecutive bowed nasal-coda words “天空”

Figure 30: Spectrogram of the spoken nasal-coda word “香”

Figure 31: Spectrogram of the bowed plosive-coda word “極”

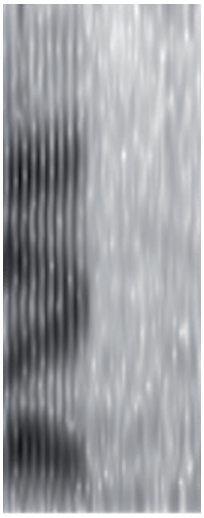


Figure 32: Spectrogram of the spoken plosive-coda word “歷”

As compelling as these visual representations are, they offer only flickers of correspondence at the microscopic level. However, zooming out to compare larger phrasal units proved unproductive, as the spectrographic signatures of wordless bowed utterances and spoken words are inherently incommensurable (see Figures 33 & 34). A forced reading risks imposing a false gestalt, connecting dots where none meaningfully exist.

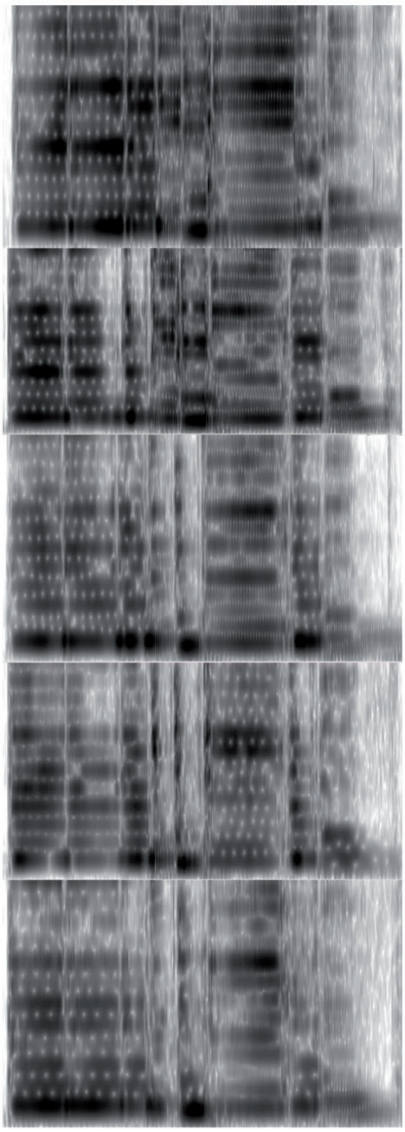


Figure 33: Spectrograms of the bowing-schemata realizations of the first phrase: Japanese, Cantonese, English, Mandarin Chinese, and Malay (top to bottom)

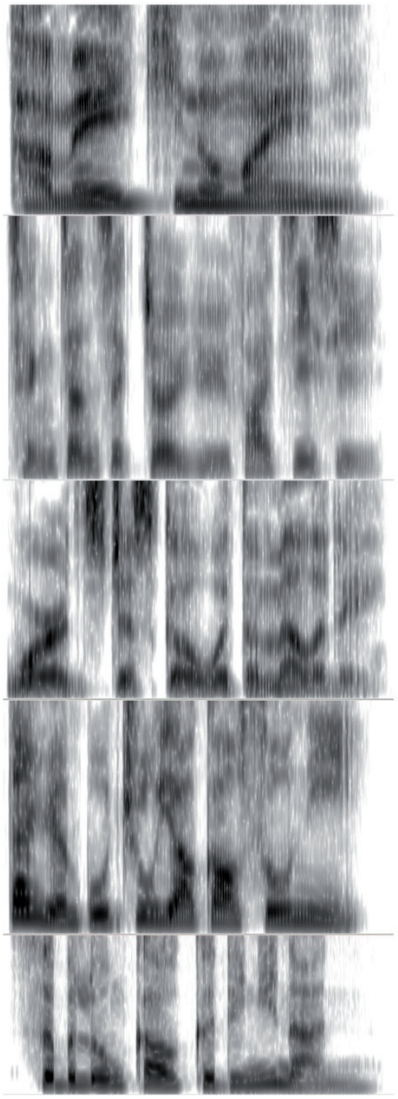


Figure 34: Spectrograms of the first verse line generated by Speechify: Japanese, Cantonese, English, Mandarin Chinese, and Malay (top to bottom)

Nevertheless, the promising glimpses of iconic native prosodic markers in cello playing – from Bylsma's Dutch guttural timbre to Gaillard's French final stress to my Cantonese syllable-final articulation – point to the plausible existence of a "linguistic aura" embodied in string performance. By absorbing and attending to the totality of words: their semantics, their syntax, and their phonetic grain, the HIP community has breathed life into *Non l'intendite parlare?*

5. Conclusion

Do we have an "accent" when playing in the same way we speak? That depends on whether the involuntary internal dialogue running in our mind is "accented." Since our native language shapes how we listen (Cutler, 2012), and any subsequent auditory input is adapted through its sonic framework cultivated from our early years, it is reasonable that musicians who dedicate their lives to sound crafting have long sensed the connection between linguistic imprint and musical realizations. In order to investigate this elusive phenomenon, I propose using the bowing schemata as a springboard for diving into the articulative nuances expressed in both string playing and speech sounds. Speech by nature is "fast, continuous, variable, and nonunique" (Cutler, 2012: 39); however, for the purpose of unraveling the mystery of auditory cognition, it is necessary to treat it as a slow, discrete, stable, and unique entity to discern its particularities. The main challenge is to define what counts as a distinct quality of a language: is the articulatory gesture effortlessly produced by a native speaker yet causing great resistance for a non-native speaker a diagnostic marker? The deliberate act of sound copying, through the use of the extra-responsive Baroque bow, opens a path to evaluate our subjective experience. This process of intentional engagement, translation, and embodiment enables a direct interrogation of "native-ness" and "foreign-ness," and the derived blueprint provides a point of departure for further empirical exploration (see Fan, 2025 for an initial study).

Identifying "native-ness" in wordless instrumental expression is understandably a sensitive topic as it entangles schools of training, aesthetic ideals, and listening exposure. If we associate a musical "accent" with stylistic literacy, we face the danger of cementing the idea that a predominant, tasteless trend exists in performance. This framing of "accent" as a marker of deviation finds its echo in Graham's (1969: 445) strongly worded interpretation of it as "speaking a language in a way that betrays the speaker's national or geographical background." However, the discomfort experienced by the sensitive ear − that reflexive tension provoked by accents in spoken language − need not be translated into the musical realm. The beauty of musical micro-expressions is founded precisely on the acceptance of diverse "accents" shaped by our native language. The bowing schemata was formulated with this in mind and in alignment with the broader HIP ideal of attending to subtle nuances to elevate musical expression. By tapping into the vast inventory of prosodic gestures from myriad languages, we can navigate fluidly between the stock of phonetic patterns we are aurally wired with and those sensuous utterances encountered in foreign languages, to enrich our musical delivery. Just as syntactic strings guide the mind, so does linguistically-informed string articulation evoke the beauty that resonates with the passions.

References

Arbib, M.A. (2013). *Language, music, and the brain: A mysterious relationship*. Cambridge, Massachusetts: The MIT Press.

Bach, J.S. (1992). Cello suite No. 3 in C major, BWV 1009: Prélude [Musical work recorded by Anner Bylsma]. On J. S. Bach: Suites for violoncello solo, BWV 1007–1012. Sony Classical. (Original work published ca. 1720)

Bach, J.S. (2011). Cello suite No. 3 in C major, BWV 1009: Praeludium [Musical work recorded by Ophélie Gaillard]. On Bach: Suites pour violoncelle seul (intégrale). Aparté. (Original work published ca. 1720)

Bach, J.S. (2020). Cello suite No. 3 in C major, BWV 1009: Prelude [Musical work recorded by Alisa Weilerstein]. On Bach: Cello suites. Pentatone. (Original work published ca. 1720)

Barbero, C., & Calzavarini, F. (2024). *Experiences of silent reading. Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-024-09966-x

Bernstein, L. (1976). *The unanswered question: Six talks at Harvard*. Cambridge, Massachusetts: Harvard University Press. (Lecture series delivered 1973)

Brandt, A., Gebrian, M., & Slevc, L.R. (2012). Music and early language acquisition. *Frontiers in Psychology, 3*, Article 327. https://doi.org/10.3389/fpsyg.2012.00327

Caccini, G. (2021). Le nuove musiche (L. Abadie, Trans.). In E. Rotem & T. Braithwaite (Eds.), *Giulio Caccini's published writings: Bilingual edition* (pp. 24–46). Early Music Sources. (Original work published 1602) https://www.earlymusicsources.com/pie#h.jcqlsz12gfcn

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, Massachusetts: The MIT Press.

Easy Dutch. (2023, June 9). 100 words you should know when coming to the Netherlands | Super easy Dutch 20 [Video]. YouTube. https://www.youtube.com/watch?v=jSyrqH_MMOM

Easy French. (2023, October 22). What do French people actually eat? | Easy French 189 [Video]. YouTube. https://www.youtube.com/watch?v=p65EBC9lW9k

Easy Languages. (2021, January 20). Hong Kong Museum of History | Easy Cantonese 8 [Video]. YouTube. https://www.youtube.com/watch?v=4qb5Oghwv-Q

Egginton, W. (2009). *The theater of truth: The ideology of (Neo)Baroque aesthetics*. Stanford, California: Stanford University Press.

Everett, P. (2002, December). George Frideric Handel: No, di voi non vo' fidarmi, HWV 189. Edition HH. https://www.editionhh.co.uk/hh40pref.htm

Fan, D. (2025). Perception of musical articulation: Does native monosyllabic or polysyllabic language determine perceived naturalness? [Poster session]. *The 18th International Conference on Music Perception and Cognition*, São Paulo, Brazil.

Faudree, P. (2012). Music, language, and texts: Sound and semiotic ethnography. *Annual Review of Anthropology*, *41*, 519–536. https://doi.org/10.1146/annurev-anthro-092611-145851

Fishman, G. (2014). On "what makes a Baroque cellist": Foreign languages: Part 2. cellobello.https://cellobello.org/cello-blog/baroque/on-what-makes-a-baroque-cellist-foreign-languages-continued/

Geeting, J. (2008). *Janos Starker: "King of cellists": The making of an artist*. Simi Valley, California: Chamber Music Plus Publishing.

Graham, R.S. (1969). The music of language and the foreign accent. *The French Review*, *42*(3), 445–451.

Hall, R.A.Jr. (1972). Elgar and the intonation of British English. In D. Bolinger (Ed.). *Intonation: Selected readings* (pp. 282–285). Harmondsworth: Penguin Books. (Reprinted from "Elgar and the intonation of British English," 1953, Gramophone, 31, 6.)

Haynes, B., & Burgess, G. (2016). *The pathetick musician: Moving an audience in the age of eloquence*. New York: Oxford University Press.

Hooker, L.M. (2013). *Redefining Hungarian music from Liszt to Bartók*. New York: Oxford University Press.

Hui, T. (2020). *Melodramas of the tongue: Accented speech in literature, art, and theory* (Doctoral dissertation). Leiden University, Leiden, Netherlands. https://hdl.handle.net/1887/136967

Keiler, A. (1978). Bernstein's "The Unanswered Question" and the problem of musical competence. *The Musical Quarterly*, *64*(2), 195–222. https://www.jstor.org/stable/741445

Kubota, T. (1996). La la la love song [Song]. On La la la love thang. Sony Music; Japan.

Novembre, G., & Keller, P.E. (2014). A conceptual review on action-perception coupling in the musicians' brain: What is it good for? *Frontiers in Human Neuroscience*, *8*, Article 603. https://doi.org/10.3389/fnhum.2014.00603

Oxford University Press. (n.d.). Accent, n. In Oxford English dictionary. Retrieved September 1, 2025, from https://doi.org/10.1093/OED/9494434378

Palmer, C., & Hutchins, S. (2006). What is musical prosody? *Psychology of Learning and Motivation*, *46*, 245–278. https://doi.org/10.1016/S0079-7421(06)46007-2

Patel, A.D. (2008). *Music, language, and the brain*. Oxford: Oxford University Press.

Patel, A.D., & Daniele, J.R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), B35–B45. https://doi.org/10.1016/S0010-0277(02)00187-7

Quantz, J.J. (1966). *On playing the flute* (E.R. Reilly, Ed. & Trans.). London: Faber and Faber. (Original work published 1752)

Reiter, W.S. (2020). *The baroque violin & viola: A fifty-lesson course* (Vol. 1). New York: Oxford University Press. https://doi.org/10.1093/oso/9780190922696.001.0001

Scharinger, M., & Wiese, R. (Eds.). (2022). *How language speaks to music: Prosody from a cross-domain perspective*. Berlin: De Gruyter.

Sherman, B.D. (1997). *Inside early music: Conversations with performers*. New York: Oxford University Press.

Steiner, G. (1975). *After Babel: Aspects of language and translation*. New York: Oxford University Press.

Taruskin, R. (2023). *Musical lives and times examined: Keynotes and clippings, 2006–2019*. Oakland, California: University of California Press.

Temperley, D. (2022). Music and language. *Annual Review of Linguistics*, *8*, 153–170. https://doi.org/10.1146/annurev-linguistics-031220-121126

Temperley, N., & Temperley, D. (2013). Stress-meter alignment in French vocal music. *The Journal of the Acoustical Society of America*, *134*(1), 520–527. https://doi.org/10.1121/1.4807566

The Juilliard School. (2016, January 7). Juilliard snapshot: Robert Mealy [Video]. YouTube. https://www.youtube.com/watch?v=hIb9ygatTC0

The Strad. (2022, April 12). Technique: Speaking with the bow. The Strad. https://www.thestrad.com/playing-hub/technique-speaking-with-the-bow/14614.article

Vanscheeuwijck, M. (2020). Violoncello and other bass violins in Baroque Italy. In D. Fabris (Ed.), *Gli esordi del violoncello: A Napoli e in Europa tra sei e settecento* (pp. 25–100). Barletta: Cafagna Editore.

GOVOREĆE GUDALO: LINGVISTIČKE REZONANCE U GUDAČKOM IZVOĐENJU

Apstrakt

Fenomen podsvesnog lingvističko-prozodijskog otiska proučavan je u muzičkoj kompoziciji (Patel & Daniele, 2003; Temperley, 2022), ali njegovo delovanje u izvođenju i dalje ostaje *terra incognita*. Ovaj jaz predstavlja prepreku za procenjivanje načina na koji izvorni fonetski obrasci prikriveno oblikuju „osećaj“ u muzičkom izrazu – idiosinkratično kodiran u artikulaciji. Kako bi se ispitala ova utelovljena interakcija, naša studija uvodi novu metodologiju koja koristi neuporedivu artikulacionu slobodu baroknog gudala da bi se destilovali jezički „ukusi“. Višejezične adaptacije pesme *La La La Love Song* Tošinobua Kubote služe kao fonetski šabloni za prevođenje prepoznatljivih prozodijskih obeležja – uključujući japanski visinski akcenat, kantonske završetke slogova i engleski ritam zasnovan na naglasku – u ponovljive šeme gudalovanja. Preliminarne analize spektrograma ukazuju na to da se artikulacija gudača usklađuje sa izvornom prozodijom, naročito u izvođenju obojenom govorom. Na primer, spektralne osobine poteza Annera Bilsme odražavaju holandske grlene timbre dok oblikovanje tonova kod Ofeli Gajar replikuje pojačavanje francuskog završnog leksičkog naglaska. Predložene šeme nude nacrt za ispitivanje „izvornosti“ i „stranosti“ u izvođenju. Mapiranjem fonetskih obrazaca na gudalačke geste, ovaj rad pruža metod za analizu prožimanja izvorne prozodije u instinktivnom sviranju, usavršavajući izražajnu nameru kroz svesnu artikulaciju senzualnosti koja je inherentna jezičkom zvuku.

Ključne reči: govoreće gudalo, muzička artikulacija, lingvistički akustički obrasci, međuigra jezika i muzike, kulturna rezonanca

THE SASE JOURNAL
ISSN 3042-2930, UDK 821.111

№ 2, 2026
pp. 79-90

Original research paper
UDC 781:81'37
DOI: https://doi.org/10.46630/sase.2.2026.2
ORCID: 0000-0002-1655-8485

IMAGE SCHEMAS IN INTERACTION BETWEEN LISTENERS AND INSTRUMENTAL MUSIC

Violetta Kostka[1]
Academy of Music in Gdańsk, Poland

Abstract

Although they are fundamental to our basic life experiences, image schemas are rarely discussed across disciplines. In fact, knowledge about them is almost non-existent within the music community. This article aims to present the modest musicological literature on the subject, explore image schemas in *Two Studies* for piano (1986) by Paweł Szymański, and examine image schemas identified in music history. *First Study*, a post-tonal composition in eight episodes, features series of identical chords of increasing length, where the first chord is always loud and subsequent chords softer. Interpreted as an expanding musical echo, it requires image schemas such as ITERATION, FORCE, PATH, and UP. *Second Study* is a fast-paced, one-voice piece alternating between quasi-baroque and modernist sections. Assuming that we consider the etude as a rapid musical movement in two different manners − one with a well-defined goal and one with a goal that resists clear definition − it incorporates the following image schemas: +PATH and +GOAL for quasi-baroque sections, +PATH and –GOAL (or TELEOLOGICAL MOVEMENT) for modernist sections, and CYCLE for the entire study. According to the author, the enumerated schemas are highly characteristic of music, regardless of genre, style, or musical system (major-minor, post-tonal, or others).

Keywords: image schema, conceptual blending, meaning, Paweł Szymański, *Two Studies*

1. Introduction

Modern research on meaning, including musical meaning, initiated several decades ago has proven very fruitful. Today, we not only have constantly evolving

[1] Email address: v.kostka@amuz.gda.pl
Corresponding address: Academy of Music, ul. Kmieca 5, 80-279 Gdańsk, Poland

subdisciplines such as image schemas, frames, conceptual metaphors, and conceptual integration, but we are also deeply and increasingly engaged with meaning stemming from our experience. This article is devoted to the problem of image schemas in the context of music. I will first present one of the most recent positions on image schemas in cognitive science, then move on to discuss this problem in music, touching on the literature, my own research results, and broader implications.

2. Image schemas in cognitive science

Image schemas are organizing anchors of cognition common to all human beings or prelinguistic ontologies concerning space, time, and other core elements of embodied human experience used to conceptualize the world. The notion was introduced to science almost half a century ago, and today the list of works devoted to it is quite extensive (Johnson, 1987, 2007, 2018; Mandler, 2004; Rohrer, 2005; Oakley, 2010; Mandler and Cánovas, 2014). Without denying the importance of earlier findings on schemas, I offer below a brief summary of the position of representatives of the neural theory of linguistics, by George Lakoff and Srini Narayanan, as presented in the book *The Neural Mind: How Brains Think* (2025: 92-112).

The authors claim that an image schema represents a general case − one that is understood cognitively but cannot be directly perceived. Immanuel Kant's example of an image schema of a triangle illustrates this distinction. "This means," they wrote, "that if you activate the [neural—V.K.] circuitry for any specific schema, you will also activate the circuitry for the general schema. But the general schema does not necessarily activate any specific schema" (2025: 94). Lakoff and Narayanan term these basic schemas as primary schemas, categorizing them into three types: image schemas, force schemas, and executing networks (X-Nets). While the first two categories are widely recognized and accepted, the third—executing networks—was introduced by the authors. For instance, to trace a triangle in the air with your index finger, you need X-schema circuitry in your brain, neurally bound to triangle schema circuitry. As noted, primary schemas are embodied, with the most definitive examples being the motion schema, the part-whole schema, the balance schema, the container schema, the contact schema, among others. These schemas play a pivotal role in conceptual thought, primarily because "they can be combined to form complex schemas, and they are used (…) in other kinds of conceptual structures: frames, conceptual metaphors, and conceptual integration" (2025: 104).

The most compelling data pertains to explanations centered on pattern-based thinking – logic schemas. According to Lakoff and Narayanan, embodied schemas underpin inferences that we unconsciously and effortlessly employ in daily life. However, the human brain does not possess specialized neural circuits for every individual inference. The theorists propose that these inferences are facilitated by image schematic generalizations, suggesting that "the special case circuitry adds additional structure to the general case in each specific case (…), that general

circuitry is physically part of each specific case of neural circuitry" (2025: 107). For instance, a single logic schema can accommodate embodied schemas such as source-path-goal, container, behind, and above. These findings lead the authors to the concept of simulation, which they argue is integral to everyday cognition. In their view, inferences or logic schemas function as essential components of mental simulation. As they say:

> For that to happen, the right combinations of neurons have to be in right places to constitute a circuit for a logic schema, and for an embodied schema that is relevant in this situation. Once recruited through repeated use, those neural circuits keep firing in appropriate situations, getting stronger and stronger until they are permanent. That is how their content is learned (2025: 113).

Lakoff and Narayanan's attempt to link elements of thought and language with neural circuitry is an intriguing, though rather isolated, direction for the development of image schemas. It is possible that this approach will inspire further in-depth studies of these fundamental units of our thought processes.

3. Image schemas in music cognition

In contrast to cognitive linguistics, the fields of systematic musicology and music theory have yet to develop a substantial body of research on image schemas. Three pivotal contributions merit attention: Candace Brower's seminal article (2000), which laid the foundation for this line of inquiry; the recent comprehensive study by Mihailo Antović, Vladimir Janović, and Vladimir Figar (2023), offering a novel perspective on the subject; and Mihailo Antović's chapter focused on the concept of the meta schema (forthcoming).

Drawing upon the basic conceptual metaphors MUSICAL EVENTS ARE ACTIONS and A MUSICAL WORK IS A JOURNEY, Brower attempted to identify image schemas characteristic of melody, harmony, phrase structure and narrative within major-minor music. She posited that a tonal melody, moving primarily through diatonic steps and reaching a final rest on the tonic (the most significant pitch), is underpinned by six schemas: SOURCE-PATH-GOAL, VERTICALITY, CONTAINER, CENTER-PERIPHERY, BALANCE, and CYCLE. For instance, the VERTICALITY schema pertains to the hierarchical significance of pitches within a given key. The tonic pitch is considered the lowest and most fundamental, while the subsequent degrees of the triad are "upward" (in frequency), and the highest is the eighth degree, which reiterates the tonic an octave higher. "As a consequence of our interpreting melodic tones as having differing degrees of stability, we experience them as acted upon by *forces*. We feel these forces to act most strongly on the unstable tones of a melody, pulling them upward or downward to the closest stable tones" (2000: 334). These forces are compared to a constant downward pull of gravity. After presenting her theory, Brower offers a nuanced semantic interpretation of Schubert's song *Du bist die Ruh*. She identifies the presence of CONTAINER

and SOURCE-PATH-GOAL schemas within the poem, and CYCLE, CONTAINER FOR MOTION, and EXPANDING CONTAINER schemas within the music. She concludes her article by asserting that the presence of image schemas and metaphors has been substantiated in major-minor classical pieces and world music, and she posits that it is highly probable they also manifest in post-tonal/neotonal compositions.

The study conducted nearly a quarter of a century later by Antović et al. (2023) confirms some of Brower's findings while introducing novel dimensions to the discourse. The authors delve into the generation of meaning in both music and language, yielding intriguing comparative insights. Their investigation centers on two hypotheses. The first posits that multiple image schemas are concurrently engaged at any given moment in the cognitive process, interacting dynamically. The second hypothesis elevates the image schema SCALE to a higher-order parameter, suggesting it serves as an indicator of the intensity of the schemas involved. The Serbian researchers propose a three-tiered scale for each schema, encompassing upward and downward gradations. Interestingly, they believe that studying complex image schemas is easier in music than in language.

Conclusions regarding music are discussed using an excerpt from Beethoven's Piano Sonata No. 1 in F minor (bars 6-7). The analysis of musical structure, the distinct climax and ornaments contained within, reveals that the musical material evokes the following image schemas in the listener: PATH, FORCE, BALANCE, LINK, and CONTAINMENT. The relationships between these schemas are dynamic, changing over time, as best demonstrated by the scale of positive or negative scores assigned to them. The authors draw the following conclusions from their study: the greater the number of active image schemas during the reception of a musical fragment, the more diverse extramusical associations the listener can generate. Furthermore, the stronger the scalarity, the more profound the emotional impact.

In the third announced work, Antović (forthcoming) updates his multilevel-grounding theory (2022) and postulates an enhanced role for cross-modal correspondences in generating semiosis. He advocates for replacing embodied image schemas with the concept of the meta-schema, defined as "a set of higher constraints, which are likely amodal, though they may receive more strongly embodied specifications at a later stage of meaning construction" (forthcoming: 2). This abstract concept aims to integrate various embodied image schemas into multilevel-grounding, each contributing distinct semantic trajectories at higher levels. As an illustrative example, the renowned piccolo motif from Mozart's *The Magic Flute* is presented here, with an emphasis on pitch movement. Antović juxtaposes open-ended descriptions of this motif provided by experiment participants with phrases derived from musicological analyses as well as with three distinctly shaped scenes featuring this motif in opera performances from recent years. The conclusion from this research is as follows:

> Instead of a variety of schemas for directed movement (…) I have proposed that the succession of pitches in musical scales is naturally conceptualized on the basis of a meta-schema, which introduces a discretely ordered, stepwise, unidirectional path or

> transformation of a geometric shape, leading from smaller to higher consumption of energy during the motion. In turn, this meta-schema may lie at the basis of numerous instances of musical meaning generation sparked by pitch succession, verbal, imagistic, or fully multimodal, as in operatic and film settings (forthcoming: 10).

It is noteworthy that Antović, at the time of writing this chapter, was not yet familiar with the work by Lakoff and Narayan discussed above, which posits that neural circuitry for specific schemas activates the circuitry for general schemas. The convergence of ideas here is likely not coincidental, warranting further investigation.

4. Schemas discovered while listening to Paweł Szymański's *Two Studies* for Piano

Following Candace Brower's suggestion to expand musicological research on image schemas in post-tonal music, I determined that the two-level music ("surconventional" music) of the leading Polish contemporary composer Paweł Szymański (b. 1954) would serve as excellent research material (Kostka, 2018a, 2018b, 2021, 2022). *Two Studies* for piano (1986) was selected as a case study. These compositions rank among Szymański's most popular works, readily available on multiple CDs[2] and online,[3] which I recommend listening to in order to follow my reflections. These works were crafted in a manner typical of the Polish composer. In each etude, the starting point was a minor structure (chordal and melodic, respectively), which was then radically expanded using some mathematical ideas and embellished. The post-tonality evident in the *Two Studies* signifies a substantial disruption of traditional tonality, stemming from the adopted two-level compositional technique.

My research commenced with formal analyses of the pieces and an exploration of the intuitive extra-musical meanings attributed to them by music critics, pianists and a few scholars. By narrowing these meanings to a single recurring theme, I sought to demonstrate that this chosen meaning is not arbitrary but is underpinned by a reasoned thought process, specifically as a result of conceptual blending (Fauconnier & Turner, 2002; Oakley & Pascual, 2017). With such a robust semantic structure that I inferred, I further aimed to identify a set of image schemas underlying the meaning generation process.

First Study (*Presto ritmico sempre staccato e secco*) originates from a baroque chord structure, which has been transformed and extended through the application

[2] *Paweł Szymański. Partita III, Lux aeterna, Partita IV, Dwie etiudy, Miserere*, Warszawa: Accord 1997, piano - Szábolcs Eszténvi; *Paweł Szymański. Works for Piano. Maciej Grzybowski. Piano*, Warszawa: EMI 2006; *Simon Ghraichy 33,* Deutsche Grammophon 2019; Paweł Szymański, Andrzej Ślązak Works for piano. Opus Series 2019.

[3] Both Studies https://www.youtube.com/watch?v=-8sKxS0irjY. First Study: https://www.youtube.com/watch?v=_z7DBSBXsd4&ab_channel=Ari. Second Study: https://www.youtube.com/watch?v=K2VkfjHvOjg.

of mathematical principles. The final composition features series of identical chords, with the first chord in each sequence being loud and subsequent chords always softer. The loud chords enter at irregular gaps, but the repetitions of each loud chord maintain consistent gaps. At the beginning of the study, series of identical chords are distinctly audible, but as the series begin to overlap, their clarity diminishes significantly. All chords follow an eighth-note rhythm in a five-eight meter, yet the study lacks a regular beat. The piece is divided into eight episodes, with the number of identical chords in each series increasing progressively from 1 to 8 across the episodes. For instance, the second episode comprises only series with two identical chords, whereas the third episode includes only series with three identical chords.

First Study, like *Second Study*, represents a virtuosic genre, which might suggest that it will be perceived by listeners as a self-referential work. However, this assumption is far from accurate. I managed to find a dozen various meanings in the texts about the piece, written, among others, by music critics (e.g. Marcin Gmys, Tomasz Cyz) and pianists (e.g. Maciej Grzybowski, Simon Ghraichy). All of them, grouped by similarity, are presented in Table 1.

Meanings attributed to *First Study*
"expanding musical echo" "echo effect" "more and more frequent reflections in the form of echoes" an étude "immersing itself in the affect of an apparent echo"
"we are witnessing a kind of game with time. The subsequent chords give the illusion of a double, triple or even quadruple reality" "a kind of disturbance in time" "the slow emergence of music"
the "self-propelling mechanism carries compressed emotional content" "maximum emotion"
"waterfalls of chords"
"ametric intrigue" gives the study "a dramatic context"
"the mystical aura of this brilliant composition"
the beginning of the work as a "broken tango"
"a hypnotic effect on the listeners"

Table 1. Meanings attributed to *First Study*

Since the first group of meanings, associating this etude with the concept of echo, is the most numerous, I delve into how this meaning takes shape. Imagine a conceptual integration network diagram appropriate for a fragment of the piece, where two input spaces converge: one rooted in the physical world of echoes, and the other in the musical realm of repeating chords. In the echo input, we find a loud sound source and its reflections, while the musical input mirrors this with a loud chord followed by softer repetitions. As a result of correspondences between the elements from both inputs and the projection of elements from the inputs to the blending space, a possible meaning emerges − a musical echo. But the story does not end there. This etude unfolds like a narrative, with seven pivotal moments where a series of identical chords meets another series, one chord longer. To capture this transition, a new type of conceptual integration network is needed. In this network,

elements from two different meaning frames correspond: in one, we have the familiar everyday experience of increasing the number of phenomena/objects (e.g., books on a shelf), expressed by numbers from one upward (1, 2, 3, 4, 5, 6, etc.); in the other, we have transition from one episode to another, in which one series intersects with another series which is longer by one chord. As these elements intertwine and project into the blended space, a deeper meaning surfaces − an expanding musical echo.

How did several listeners arrive at nearly identical interpretations of Szymański's *First Study*? The answer lies in their unconscious reliance on similar image schemas. Consider one series of identical chords from any episode. Listening to such a series activates the ITERATION image schema (a thing or process is repeated a certain number of times), which is closely tied to the conceptualization of sound source repetitions in both natural phenomena and music. The next image schema at play is FORCE, which here pertains to the volume of the sound. In both echo and the piece, volume decreases over time − earlier sound is louder, later sounds softer. Since a series of identical chords unfolds over time, the PATH image schema becomes relevant, and during the transition from one episode to the other, the UP image schema emerges. Given the piece's relative uniformity in technical means, the set of image schemas remains consistent throughout *First Study*.

Second Study (*Prestissimo senza metro ma ritmico*) emerges from tonal, baroque sequential writing, which Szymański has transformed, expanded, and embellished. The resulting form of this piece is highly unconventional. It consists of a single-voice melody, flowing continuously in sixteenth notes without a discernible meter or beat. The composition alternates between approximately 70 quasi-baroque sections and 70 modernist sections. Initially, these sections are very brief, but they gradually lengthen as the piece progresses. The distinction between these section types becomes particularly evident in the longer sections.

As with *First Study*, *Second Study* has evoked a broad range of meanings among music writers (e.g. Andrzej Chłopecki, Maja Trochimczyk, Ewa Szczecińska, Maciej Grzybowski). These interpretations encompass themes of time, emotions, transcendence and many others (see Table 2).

Meanings attributed to *Second Study*
"the titanic flows of sounds"
"the idea of a single-voice fugue" [fugue = an escape]
a study "looking for a melody, only to lose it and find it"
"a rapid musical movement in two alternating manners: one with a well-defined goal and one without a goal"
"time stops, although paradoxically there is intense 'happening'" "a kind of disturbance in time"
"there is a constant pulsing energy causing it to spin lines around its own axis"
the study is an „arabesque"
the study is „crystalline", "passages like glass beads"
"everything directed all the time to a higher space, towards the sun, the sky, or maybe God"
"a sense of reverberation", like "in some chapel from old days"
"maximum emotion"
"a hypnotic effect on the listeners"

Table 2. Meanings attributed to *Second Study*

In this case, it seems that there is only one group with similar meanings (concerning time), but the first four are still strongly interconnected. The first two meanings clearly pertain to movement, the third poetically captures the alternation between section types, while the fourth unites these earlier ideas together. Let us consider the fourth meaning. While listening to any longer quasi-baroque section, two mental spaces participate in its interpretation: physical movement and musical event. The physical movement encompasses elements such as an agent in space, regular and rapid motion, change of location, and a clear destination for the movement. In turn, the musical event comprises a melody with doubling of each note, regular motion at a fast tempo, diverse motifs, and gravity towards the most significant pitch − the tonic. From the correspondence of these elements and their projection into the blending space, the section's meaning emerges − a musical movement akin to a fast run with a precisely defined goal. In the case of any longer modernist section, many elements remain consistent, yet there are also distinctly different elements, such as a melody composed of numerous indirect repetitions of individual notes and the absence of gravity. Such a melody evokes associations with rapid physical movement, like pounding on an electric treadmill. Consequently, I define the section by its meaning − a musical movement resembling a fast run without a defined goal.

All the meanings created for *Second Study* are the result of the unconscious participation of image schemas. Two fundamental image schemas come to the fore: PATH and GOAL. Lakoff and Johnson proposed a combined version: SOURCE-PATH-GOAL, which includes: a trajector that moves, the starting point, a goal representing an intended destination of the trajectory, a route from the source to the goal, and additional elements (1999: 33). However, I will use PATH and GOAL separately, whereby PATH signifies the trajectory or route of movement, and GOAL denotes reaching an intended destination. With this assumption, the quasi-baroque section draws on a pair of image schemas: +PATH and +GOAL, while the modernist section relies on the pair: +PATH and −GOAL, where the minus sign indicates that the goal remains elusive due to the intricate structure. Instead of –GOAL, the thought process focused on the modernist section could draw on a different schema, such as TELEOLOGICAL MOVEMENT understood as "moving toward a locationally unspecified 'goal'" (Antović, forthcoming: 2). Furthermore, the CYCLE image schema appears quite evident in this etude. According to experts, CYCLE is used whenever a series of events occur in a specific order and are often repeated. In the case of *Second Study*, one cycle comprises one quasi-baroque section and one modernist section, and there are approximately 70 such cycles in total. A listener experiencing this etude from beginning to end cannot fail to notice this.

5. Broader implications

The image schemas enumerated above, connected with Szymański's post-tonal *Two Studies*, are not uncommon in the experience of listening to musical pieces. It is

conceivable that they might be activated regardless of genre, style, or system (major-minor, post-tonal, or any other). Below, I provide several examples.

ITERATION is likely a fundamental organizing anchor of music cognition, as repetition itself is a significant compositional element. Various forms of repetition are evident in Stravinsky's music (Horlacher, 2011) and in the genre of canon, known since the Middle Ages. However, the highest frequency of repetition is found in minimal music. For instance, in Terry Riley's *In C*, there are 53 melodic-rhythmic cells, each repeated until the conductor signals a transition to the next cell. Repetition also serves as a crucial foundation for ritual forms (e.g., Zikr) and in the world music repertoire. Its participatory nature can draw listeners into an immersive relationship with music, laying the groundwork for diverse affective and meaningful responses (Margulis, 2013).

The UP image schema and its opposition DOWN are fundamental to music, as composers have employed expansion and contraction techniques for centuries. These techniques are extensively utilized by Arvo Pärt (Shvets, 2014) and Rafał Augustyn (Ferenc, 2024) to shape their musical forms. Polish musicologist Maciej Gołąb has examined such phenomena in music, categorizing them as telescopic and chipped techniques. The former has been identified and described in the works of Beethoven, Chopin, Mahler, and Bartók, while the latter is evident in Karol Szymanowski's *Pentezilea* (Gołąb, 2011: 78−80).

Another image schema identified in my interpretation above is CYCLE, which may refer to several distinct music phenomena. In acoustic terms, a cycle denotes a single complete and recurring sequence of compression and rarefaction in air pressure. In musical composition, a recurring rhythmic pattern can also exemplify a cycle. Numerous musical traditions have been characterized as fundamentally cyclical in nature. For instance, in the gamelan music of Indonesia, nested gong cycles − known as colotomy − structure the rhythmic framework of a piece (Becker, 1984; Tenzer, 2000). Other recurring rhythmic structures can be found in Indian classical music; they are known as tala.

Many scholars concur that music's temporal development aligns with the SOURCE-PATH-GOAL image schema. As Brower observes, this schema is most prominently exemplified in the major-minor tonal system, which dominated Western music from the 17th to the 19th century. This dominance is largely attributed to the structural inclination of major-minor music to resolve into a cadence centered on the tonic.

In light of compositional experiments of the twentieth and twenty-first centuries, in which the obvious tonal-harmonic trajectory toward a final tonic has become either secondary or altogether irrelevant, the conceptual integrity of the SOURCE-PATH-GOAL image schema has been notably disrupted. For instance, Michael Spitzer identifies a lack of clear directional motion in Claude Debussy's *La mer* as well as in Renaissance polyphony (2022: 110). This observation may be extended to works such as Arnold Schönberg's *Farben*, which seeks to articulate a melody of shifting timbres (*Klangfarbenmelodie*); György Ligeti's *Lontano*, described by one reviewer as "a glowing lamp, to be extinguished at the end" (Willson, 2007: 114);

Louis Andriessen's *De Tijd*, which is virtually devoid of movement; John Zorn's *Road Runner*, a dense collage composed of brief excerpts from a wide array of musical sources; and Paweł Szymański's *Through the Looking Glass I*, characterized by a technique reminiscent of stuttering: following a tonal fragment, the music seems to recede, only to be followed by another tonal fragment, and so on. One may hypothesize that modern and postmodern music, like its predecessors, continues to rely on the PATH image schema, though the GOAL is frequently obscured and demands extended reflection. Regrettably, Western music education has thus far failed to encourage listeners to independently infer the GOAL of a composition − a pedagogical oversight that now appears increasingly problematic. Twentieth- and twenty-first-century music, as repertoire not yet fully explored in terms of its GOAL schema, therefore presents a compelling field of inquiry for music cognition scholars.

6. Conclusion

To summarize, the fields of systematic musicology and music theory have not yet established a robust tradition of research on image schemas. However, the efforts of certain scholars are beginning to yield promising results. Gradually, we are uncovering not only the diversity of image schemas that underpin our conceptualization of music but also refining the methodologies for employing them in subsequent research and interpretative analyses of musical works. In this article, I have sought to demonstrate that image schemas are intrinsically connected to original post-tonal music. It turned out that due to the exceptional coherence of the musical material, Szymański's *Two Studies* elicit many various meanings, and one specific meaning is supported by several image schemas. I have also provided a set of broader implications. It follows that image schemas such as ITERATION, UP, DOWN, CYCLE, PATH, and GOAL function at the interface between the recipient and the musical work, irrespective of genre, style, or system. As noted earlier, however, this issue remains relatively underexplored within the field of music cognition and calls for extensive further research.

References

Antović, M. (2022). *Multilevel Grounding: A Theory of Musical Meaning*. London: Routledge.

Antović, M. [unpublished typescript] From Embodied Imagery to Meta-Schemas: Cross-Modal Variations Provide Different, Yet Related Interpretations of the Piccolo Motive in *The Magic Flute.* The chapter intended for the edited volume entitled *The Meaning of Music: A Cognitive Approach* is currently held in the archive of Violetta Kostka.

Antović, M., Jovanović, V.Ž., & Figar, V. (2023). Dynamic Schematic Complexes: Image Schema Interaction in Music and Language Cognition Reveals a Potential for Computational Affect Detection. *Pragmatics & Cognition, 2*, 258−295.

Becker, J. (Ed.). (1984). *Karawitan: Source Readings in Javanese Gamelan and Vocal Music* (Vol. 1). Ann Arbor, Mich.: University of Michigan Center for South and Southeast Asian Studies.

Beckles Willson, R. (2007). *Ligeti, Kurtág, and Hungarian Music during the Cold War*. Cambridge: Cambridge University Press.

Brower, C. (2000). A Cognitive Theory of Musical Meaning. *Journal of Music Theory*, *44*(2), 323–374. doi: 10.2307/3090681

Ferenc, A. (2024). *Twórczość Rafała Augustyna w perspektywie intertekstualnej* (Unpublished doctoral dissertation). Kraków: Academy of Music.

Fauconnier, G., & Turner, M. (2002). *The Way We Think: Conceptual Blending and The Mind's Hidden Complexities*. New York: Basic Books.

Gołąb, M. (2011). *Muzyczna moderna w XX wieku*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.

Horlacher, G. (2011). *Building Blocks. Repetition and Continuity in Stravinsky*. Oxford: Oxford University Press.

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

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

Johnson, M. (2018). *The Aesthetics of Meaning and Thought: The Bodily Roots of Philosophy, Science, Morality and Art*. Chicago/London: The University of Chicago Press.

Kostka, V. (2018a). *Muzyka Pawła Szymańskiego w świetle poetyki intertekstualnej postmodernizmu*. Kraków: Musica Iagellonica.

Kostka, V. (2018b). Intertextuality in the Music of our Time: Paweł Szymański's Riddles. *Tempo: A Quarterly Review of New Music*, *286* (72), 42–52. doi: 10.1017/S0040298218000347

Kostka, V. (2021). Intertextual Poetics: from Ryszard Nycz's Theory to Paweł Szymański's Music. In V. Kostka, P. Castro, & W. Everett (Eds.), *Intertextuality in Music: Dialogic Composition* (pp. 87–102). London: Routledge.

Kostka, V. (2022). Paweł Szymański and His Transformation of Musical Conventions. In D. Hurwitz, & P. Eslava (Eds.), *Music in the Disruptive Era* (pp. 161–174). Turnhout: Brepols.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh. The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Lakoff, G., & Narayanan, S. (2025). *The Neural Mind: How Brains Think*. Chicago/London: The University of Chicago Press.

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

Mandler, J., & Cánovas, C. (2014). On defining image schemas. *Language and Cognition*, *6*(4), 510–532, doi: 10.1017/langcog.2014.14

Margulis, E. (2013). *On Repeat: How Music Plays the Mind*. Oxford: Oxford University Press.

Oakley, T. (2010). Image schemas. In D. Geeraerts, & H. Cuyckens (Eds.), *Handbook of Cognitive Linguistics* (pp. 214–235). Oxford: Oxford University Press. https://www.academia.edu/357340/Image_Schemas (accessed August 22, 2025)

Oakley, T., & Pascual E. (2017). Conceptual Blending Theory. In B. Dancygier (Ed.), *The Cambridge Handbook of Cognitive Linguistics* (pp. 423 – 448). Cambridge: Cambridge University Press.

Rohrer, T. (2005). Image Schemata in the Brain. In B. Hampe, & J. Grady (Eds.), *From Perception to Meaning: Image Schemas in Cognitive Linguistics* (pp. 165–196). Berlin: Mouton de Gruyter.

Spitzer, M. (2022). *Musical human*. London: Bloomsbury Publishing.

Shvets, A. (2014). Mathematical Bases of the Form Construction in Arvo Pärt's Music. *Lietuvos muzikologija*, *15*, 88-101.

Tenzer, M. (2000). *Gamelan Gong Kebyar: the Art of Twentieth-century Balinese Music*. Chicago: University of Chicago Press.

SLIKOVNE SHEME U INTERAKCIJI IZMEĐU SLUŠALACA I INSTRUMENTALNE MUZIKE

Apstrakt

Iako su fundamentalne za naša osnovna životna iskustva, slikovne sheme se retko razmatraju u različitim disciplinama. Zapravo, znanje o njima je gotovo nepostojeće u muzičkoj zajednici. Ovaj članak ima za cilj da predstavi skromnu muzikološku literaturu na ovu temu, istraži slikovne sheme u delu *Two Studies for piano* (1986) Pavela Šimanskog i ispita slikovne sheme identifikovane u istoriji muzike. Prva studija, posttonalna kompozicija u osam epizoda, sadrži nizove identičnih akorda sve veće dužine, pri čemu je prvi akord uvek glasan, a naredni akordi tiši. Tumačena kao muzički eho koji se širi, ona zahteva slikovne sheme kao što su ITERACIJA, SILA, PUTANJA i GORE. Druga studija je brza, jednolinijska kompozicija koja se smenjuje između kvazibaroknih i modernističkih odeljaka. Ako pretpostavimo da etidu posmatramo kao brzo muzičko kretanje na dva različita načina — jedno sa jasno definisanim ciljem i drugo sa ciljem koji odoleva jasnoj definiciji — ona uključuje sledeće slikovne sheme: +PUTANJA i +CILJ za kvazibarokne odeljke, +PUTANJA i –CILJ (ili TELEOLOŠKO KRETANJE) za modernističke odeljke, kao i CIKLUS za čitavu studiju. Prema autoru, navedene sheme su izrazito karakteristične za muziku, bez obzira na žanr, stil ili muzički sistem (dur-mol, posttonalni ili neki drugi).

Ključne reči: slikovna shema, konceptualno slivanje, značenje, Pavel Šimanski, *Two Studies*

THE SASE JOURNAL
ISSN 3042-2930, UDK 821.111

№ 2, 2026
pp. 91-110

Original research paper
UDC 781.21:81'22
DOI: https://doi.org/10.46630/sase.2.2026.3
ORCID: 0009-0002-9102-4691

THEMATIC AMBIGUITY AND RHETORICAL DISPLACEMENT IN MAHLER'S *FIFTH*: AN INTROVERSIVE SEMIOTIC ANALYSIS OF THE LANGSAM THEME FORMAL FUNCTION IN THE SCHERZO MOVEMENT

Zhuo Zhao[1]
Rutgers University, USA

Abstract

In the Scherzo movement of the *Fifth Symphony*, Mahler employs innovative compositional techniques that appear to diverge from Classical tradition. However, through a semiotic lens, these variations can be seen as modern extensions of Classical norms within the formal structure. Utilizing Kofi Agawu's introversive semiotic approach (2009), which applies language models in music analysis, this paper explores how the Scherzo movement's form adheres to the Classical rhetorical paradigm of Beginning-Middle-Ending (BME). Interpreting each formal section within the BME paradigm allows one to access the functional and hierarchical relationships between implicit, loosely connected events that still preserve the logic within the Classical ABA ternary form typical of a Scherzo.

Central to this movement is the thematic ambiguity surrounding the recurrence of the Langsam theme, which contrasts with the Scherzo themes and challenges conventional sectional distribution. Traditionally analyzed as the start of a new Trio section, the Langsam theme's initial appearance is accompanied by rhetorical signs that do not indicate a new section. Instead, a long Middle sign followed by a long Ending sign suggests thematic continuity rather than division. This thematic vs. rhetorical displacement leads to a reinterpretation of the traditional formal analysis as a simplified three-part form. The BME paradigm reflects a bottom-up approach, letting the traditional ternary form emerge out of the music rather than forcing the music into a preconceived mold. This semiotic interpretation of the Classical BME paradigm thus has the capacity to explain complex modern musical thoughts throughout the movement and the entire *Fifth Symphony*.

Keywords: Mahler's *Fifth Symphony*, Form, Introversive Semiosis, Beginning-Middle-Ending Paradigm, Displacement

[1] Email address: zhuo.zhao@rutgers.edu
Corresponding address: Rutgers University, 120 Neilson St Apt 429, New Brunswick, NJ 08901, USA

1. General Statement and State of Research

Although many composers went to extreme lengths to advance music to new places at the beginning of the twentieth century, they did not completely abandon their predecessors. For instance, Mahler's long symphonies exhibit revolutionary compositional techniques and complicated musical structure, reflecting many dissimilarities to the Classical and early Romantic tradition in terms of key areas, sectional division, organization of motivic and thematic materials, the length of the music, and so forth. However, rather than interpreting these dissimilarities as manifesting a new norm, I see them as deviations from the Classical norms, imbued with Mahler's own Modernist extension of formal structure and tonality. In other words, we can still interpret and analyze Mahler's symphonies according to eighteenth-century Classical standards.

This paper argues that the third movement of Mahler's *Fifth Symphony* can be viewed as a Modernist extension of the Classical norms in terms of formal structure through an introversive semiotic lens – Beginning-Middle-Ending (BME) rhetorical paradigm which combines Johan Mattheson, Schenker, and Ratner's idea for eighteenth-century music syntax model. Before interpreting the rhetorical formal features of the third movement of Mahler's *Fifth*, several statements regarding Mahler and formal studies research are necessary to discuss. In order to identify Mahler's innovation and extension of the traditional formal structure, it is necessary to compare Mahler's compositional practice with the inherited norm. Seth Monahan employs a narrative analysis of Mahler's symphonies in his book *Mahler's Symphonic Sonatas* (2015). Monahan treats sonata form as a standard storyline to which he compares individual formally idiosyncratic movements by Mahler. He also makes new interpretations of the standard formal terms from the narrative perspective. For a large work, it is necessary to apply traditional formal analysis to the organization of complex themes, sections, and movements because the musical narrative analysis by itself has no mechanism for engaging the explicit and objective logic of the music's structure. However, the musical narrative analysis includes programmatic thinking, which embraces his subjective reading to the flow of the multi-movement music. The reading is able to provide a reasonable interpretation of those ambiguous parts when interpreting with traditional formal analysis. The result of this combination of traditional formal analysis and narrative analysis is the interpretation of some of Mahler's symphonic movements as incomplete or "failed" realizations of a classical norm.

In addition, Monahan suggests the possibility of applying sonata form norms to Mahler's symphonies. However, there are issues that should be considered at this point. First, not every Mahler symphonic movement is in sonata form. In Mahler's *Fifth Symphony*, only the first two movements are in sonata form. The other three movements are in ternary or rondo form. The formal analysis of the entire symphony has much more potential to be explored besides the two sonata movements. Second, Monahan claims that the nontraditional formal phenomenon is explained as Mahler's

"deformation". However, Monahan's strategy of interpretation of Mahler's technique does not reveal the internal organization of musical ideas. Rather, he considers the dialogue between these ideas and the traditional form as an analogy to characters in a novel. Third, Monahan applies Hepokoski and Darcy's sonata theory (2006) as a principle, comparing individual work and the historical normative sonata structure, which also applies to sonata movements. In other words, Monahan's analysis focuses on the comparison between the norm and the irregular phenomenon, which leaves space for further discussion on traditional forms other than sonata form.

Besides Monahan's study about Mahler, the general formal studies create a stronger connection and potential application to the Beginning-Middle-Ending (BME) analytical approach. Caplin's article "The Classical Cadence: Conceptions and Misconceptions" (2004) focuses on the sectional divider and discusses phrases before and after a cadence. Related to the Beginning-Middle-Ending paradigm conception, Caplin's cadential progression is highly constrained, functioning as an ending group, and closely connects to the preceding material. The phrase before the ending function can be subdivided into a beginning group and a middle group.

Applying Classical formal norms to the analysis of later pieces also requires attention to the micro level. William Caplin's *Classical Form* (2000) provides a step-by-step process to look deep into a piece from small dimensions to large and from simple to complex. The theories regarding the formal structure in Classical instrumental music have precise categories, strengthening and complementing the *Formenlehre* tradition. The book exemplifies certain formal archetypes from high-Classical masters. This is valuable in the study of the formal structure of late music because the theory and principles have limitations and unambiguous formal distinctions. By comparing the standard formal structure on a small scale with a later piece, one can find which part in the music follows the inherited standard and which part deviates. Musical elements and materials which cannot be explained under the umbrella of the norms are considered idiosyncrasies. To give the idiosyncrasies an interpretation becomes the further work of many scholars.

As the implicit essence resides in the BME paradigmatic analysis, the new reading of musical events blurs the clear punctuations of certain moments, especially cadences. This means cadences feature extended or prolonged musical events to balance the massive musical material. This reminds me that the method of sectional division is also blurred in Janet Schmalfeldt's (2011:3) analytical perspective on form in early nineteenth-century music (Schmalfeldt, 2011). Schmalfeldt offers a dynamic approach to the music formal analysis and her book challenges traditional static conceptions of form by presenting an interpretive model that emphasizes process and retrospective understanding. Schmalfeldt's work is particularly relevant to the work of early nineteenth-century music, but its implications extend to a wide range of analytical and philosophical discussions on musical form.

At the heart of Schmalfeldt's study is the concept of becoming, influenced by Hegelian dialectics and modern *Formenlehre*. Rather than treating musical form as a rigid structural framework, she argues that form is something that emerges and develops dynamically through time. A passage's function is not always evident at first

but may only be understood retrospectively – what it becomes depends on subsequent events in the musical discourse. Schmalfeldt (2011:15) critiques traditional formal models, such as those of Schenker and William Caplin, while expanding upon their insights to develop a more flexible and context-based approach.

Regarding the BME analytical approach, both BME introversive semiotic approach and Schmalfeldt's context-based approach emphasize temporality and interpretative flexibility in musical form. Both approaches foreground the unfolding of musical form over time. The BME approach assigns musical passages with the Beginning function, Middle function, and Ending function, while Schmalfeldt focuses on how formal functions become what they are through temporal development. Also, both frameworks allow for ambiguity in musical signification and function. They resist strict formal definitions, promoting interpretive openness rather than fixed categorization. While they share some similarities, the BME approach focuses on pure musical rhetorical signs, which embrace music processes that represent Beginning, Middle, and Ending functions. Schmalfeldt's approach, on the other hand, emphasizes philosophical notions of process and becoming, which are different from introversive process of semiotic signs.

We have seen that scholars believe that the traditional Classical form provides Mahler a large-scale sectional direction, but the surface-level details show his deviation from the Classical norm. The question here is why are we using sonata form and other Classical norms to analyze late music? In other words, why is Mahler considered to follow the standard even though his symphonies contain so many details that make him a modern composer? To answer the question, it is worth examining his music through a different lens, the semiotic lens.

Jean-Jacques Nattiez's *Music and Discourse: Toward a Semiology of Music* (1990) encompasses both the theory of signs and the theory of meaning. The creative perspective of Nattiez's interpretation of a system of signs and meaning is that how to apprehend the meaning of the sign is determined by an individual's position in musical activity. This relates to the concept of the semiological "program". Regarding this concept, Nattiez mentions three objects: the poietic process, the esthetic process, and the material reality of the work. Correspondingly, there are three analyses: poetic analysis, aesthetic analysis, and analysis of the work's immanent configurations, which focus on understanding symbolic features. Here, the analysis of a work's immanent configurations can also be referred to as the analysis of the neutral level. It is a strength of this approach that this three-perspective scheme takes into account the composer's intentions, the pure musical materials, and the listeners' feelings, forming a complex system of musical meaning analysis.

In addition to Nattiez, Chomskian grammar (Chomsky, 1965, 2002) proposes that the ability to acquire and use language is innate to humans. It emphasizes the cognitive structures underlying speakers' implicit knowledge of their language, such as rules, syntax, and structure, rather than external behavior or mere communication patterns. To be specific, in the current Chomskian framework, syntactic structure is conceived as being built bottom-up, rather than top-down as in the earlier models (Chomsky, 1995). This view, often described as a strongly lexicalist account of

structure-building, holds that syntax is projected from the lexicon (Ninio, 2006). Consequently, there is no longer a need for abstract phrase-structure rules to generate syntactic configurations. Instead, structure emerges through the combination of lexically-based features. This shift represents a move away from abstract syntactic schemata toward a cognitively grounded model of syntax as a self-organizing system driven by lexical semantics.

Parallel to music language, this lexicalist and bottom-up conception of syntactic generation resonates with the semiotic model of musical form defined specifically by implicit beginning, middle, and ending signs. Meanwhile, the Schenkerian-influenced hierarchical thinking of surface-deep levels of BME paradigm has a similar effect to the lexical semantics in the music formal understanding. The surface realization unfolds from deep narrative potential through the self-organization of locally pure signs, which are similar to the deep structure of language, encoding meaning and logical relations.

As the topic of the paper is based on the neutral level of musical language, an ideal analytical approach is the one considering pure signs without outside references. So, I investigate introversive semiosis in the form of the Beginning-Middle-Ending paradigm without the theory of extraversive semiosis in terms of topic theory. This method is comprehensively discussed in Kofi Agawu's *Playing with Signs: A Semiotic Interpretation of Classic Music* (1991: 51). The neutral level is emphasized by his claim that "There is another class consisting of what we might call 'pure' signs, signs that provide important clues to the musical organization through conventional use, but not necessarily by referential or extramusical association" (Agawu, 1991). Agawu (1991:51) argues that "there are specific attitudes to a work's beginning, its middle, and its ending and that these strategies are an important clue to the dramatic character of Classical music."

Combining Agawu's and Caplin's terminology, I will claim that traditional formal functions become implicit via Beginning-Middle-Ending paradigmatic interpretation. The thematic concepts of phrase structure, cadence type, and position can be loosely explained as a series of events in a logical order. The value of this paradigmatic strategy is that it creates an analytical space for logically dealing with the atypical musical materials found in extended sonata form, ternary form, rondo form, and even a traditional form with polyphonic texture. So, it is an ideal method to explain Mahler's conversation between his surface-level treatment of music elements and the longer flow of a traditional form. The method contributes to the musical discourse that allows later composers, scholars, and musicians to reconsider the role of musical elements and provide references for composing, describing, analyzing, and interpreting music as a language.

2. Approach: Introversive semiotic Beginning-Middle-Ending at hierarchical levels

My analytical strategy for the study of formal complexity in Mahler's *Fifth Symphony* is to evoke the rhetorical Beginning-Middle-Ending paradigm at multiple hierarchical levels. This strategy investigates the relationship between pure musical

temporal signs and form, providing the discourse emerging from the musical materials themselves. My analytical perspective does not represent a subjective response to sonic qualities, existing outside the music itself, as narrative and topical approaches do.

Taking a semiotic analytical approach to musical form, the BME paradigmatic analysis has been divided into three hierarchical levels. Each section within this BME framework fulfills specific roles, establishing a hierarchy to show the functional relationships among segments. I suggest that Mahler's surface-level (level 1) of Beginning-Middle-Ending hierarchical signs in his *Fifth Symphony* create the impression of Mahler's "irregular" treatment of the form. The deeper levels of the BME hierarchy show the passage representing the BME function as extended events. The hierarchical thinking may raise a question about the similarity between the BME paradigmatic approach and Schenkerian analysis. Agawu (1991: 51) claims that:

> "Rather than extract various Ursätze from the pieces to be analyzed, however, I shall concentrate on the rhetorical strategy enshrined in the Ursatz. This means framing the discussion in terms of a beginning-middle-ending paradigm, the argument being that there are specific attitudes to a work's beginning, its middle, and its ending, and that these strategies are an important clue to the dramatic character of Classic music."

Agawu (1991; 56) also combines Mattheson's form of a string of verbal symbols, Schenker's two-voice contrapuntal structure, and Ratner's melodic-cadential progression together contribute to the BME model.

Derived from this conjunction, I consider the music events, covering passages of music, are represented primarily by a series of harmonies, but also melodic contour and texture, showing the same or similar specific BME function. Thus, the Beginning sign is characterized by an introductory phrase, motives, and tonic harmony, while the Middle extends and develops motives against a backdrop of predominant or sequential harmonies, characterized by harmonic instability, less thematic statement, and repetition. What I emphasize here is the Middle sign's absence of an establishment function like a Beginning sign, and a closure function like an Ending sign. Another important feature of a Middle sign is a neutral space for progressions such as chromatic development, the preparation of the tonic return, and sequences. The Ending sign consists of cadential motion in a longer passage, resolving from dominant to tonic, in most cases, also supported by a dominant pedal or tonic pedal, providing structural landmarks. The melodic, rhythmic, and dynamic perspectives contribute to the ending sign with a strong sense of closing off and completion. This approach suggests traditional formal functions are implicitly understood through the BME paradigm, presenting a series of events rather than adhering to strict sectional conventions. At its surface, this paradigm navigates the continuity and complexity of musical material, whereas at a deeper level, it upholds the logical sequence and structure, mirroring the traditional Classical form.

This approach also suits the *Fifth Symphony*'s third movement (Scherzo), where the complexity of his musical language necessitates innovative methods to balance traditional forms with substantial material. Vera Micznik (1994:119), claims

in the article "Mahler and 'The Power of Genre,'" about Mahlerian scherzo dances that: "in many Mahlerian scherzi the dances rebelliously refuse to comply with the traditional labels they have been assigned. Yet, as our example shows, while Mahler's materials defy absolute generic commitments, their behavior is dependent on the recognition of their generic models" (Micznik, 1994). Micznik highlights the paradox in Mahler's use of traditional genres, where he adheres to the genre's framework while deviating from its formal traditions. In other words, we might hypothesize that Mahler presents the scherzo movement as a ternary ABA form, with repetitions transforming it into an ABABA in a redefined pattern. Central to this Scherzo movement is the thematic ambiguity surrounding the recurrence of the Langsam theme, which contrasts with the Scherzo and Waltz themes and challenges conventional sectional distribution. The length of the third movement, spanning 819 measures, immediately impresses upon the listener a vast and complex musical context. In fact, Mahler goes beyond merely bringing more than simply enlarging a certain moment or certain cadences into an event; he also creates a displacement between thematic statements and rhetorical structure. In addition, the thematic statements and fragments frequently repeat, overlap, and hybridize, providing space for different rhetorical signs to emerge and develop. Thus, this displacement reveals the relationship between themes and rhetorical signs and, more importantly, offers a new perspective on analyzing Scherzo's three-part ABA ternary form.

3. A Scherzo section: Conventional formal analysis aligns rhetorical structure, no displacement

Before showing Langsam theme analyses, I will briefly mention the movement's initial horn call theme to show a BME rhetorical analysis. The initial theme that introduces the A section of the Scherzo, which I refer to as the "horn call" theme, cooperates with the rhetorical signs to establish clear phrases with strong Beginning signs. The horn call theme, as the Scherzo A section's opening theme, is reinforced by a "correct" supporting Beginning sign, showing no displacement. Examining the first phrase (Figure 1), played by two groups of different instruments one after another (the horns play mm. 1-8, and the woodwinds play 9-15), one will find the first clue for answering the above question in terms of the clear phrase statement with a strong indication of a Beginning sign. First of all, the opening of the phrase, or even the opening of the entire movement, is set up on the tonic triad arpeggiation in D major, indicating the tonic key in the most confirmed way. Second, the melodic line of woodwinds frequently repeats scale motion, either ascending or descending, especially at the end of the phrase, pushing to a strong PAC through an ascending D major scale. Third, the harmonic progression is very simple. Only I – V – I – V7 – I, appears as the harmonic motion underneath. In a broader sense, it is simply a I – V – I progression for the first 15 measures. Thus, the opening of the movement projects a clear and strong beginning with tonic triad arpeggiation, scale melodic motion, and I – V – I harmonies. I will also describe the opening phrase as a horn call leading

theme and this theme, as the unfolding of the structure, is a part of the indication of a deeper-level Beginning sign. Corresponding to the formal structure, this Beginning sign represents the start of the A section, the Scherzo section. Now, my focus turns to the next part of the analysis: where is the end of the A section and how does the rest form a rhetorical relationship to define the entire A section?

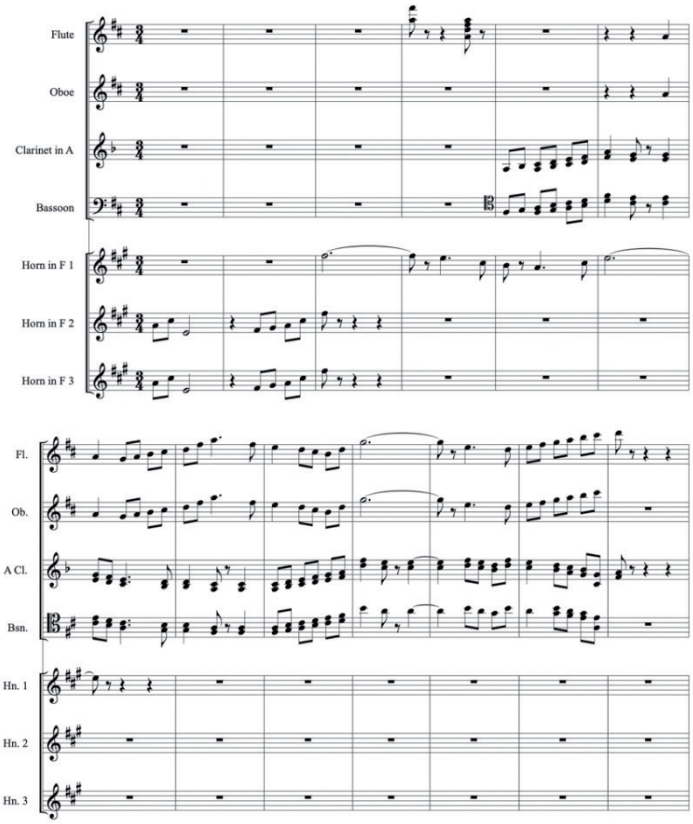


Figure 1: Mahler's *Fifth Symphony*, mvt. 3, mm. 1-15, reduced score

As the entire third movement's incredible length, it is possible to expect the large A (Scherzo) section at the beginning to be a binary form or simple ternary form. After the first surface-level Beginning sign, the following phrase repeats the Beginning sign with the same harmonic progression of I – V7 – I in mm. 16-26 with a PAC in measure 26. Measure 26 also initiates another repetition of the opening surface-level Beginning sign. In mm. 26-39, the opening Beginning sign's thematic idea is repeated in an embellished way. The harmonic structure also enlarges the basic I – V7 – I structure by adding predominants such as the vi chord and ii chord, and measure 39 closes the phrase with an IAC in D major. In summary, mm. 1-39 states the same thematic idea three times, as well as the surface-level Beginning sign three times. Thus, the first 39 measures as one group in the same thematic idea in D major, and the continuous surface Beginning sign emerges as one section of the

simple ternary's "a" section. Up to this point, a clear and confirmed Beginning sign has been established and the thematic beginning aligns with the rhetorical beginning, with no displacement (Figure 2).

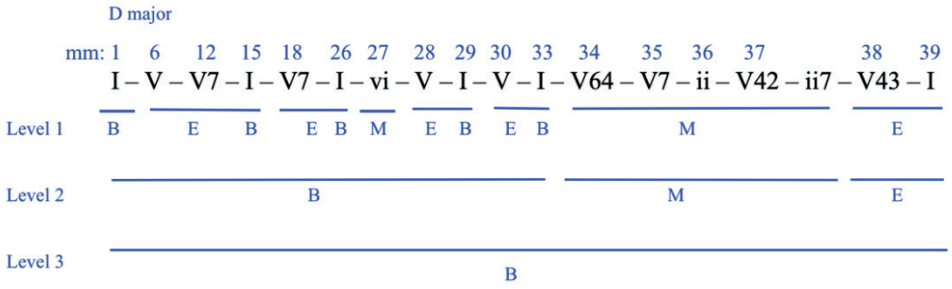


Figure 2a: Mahler's *Fifth Symphony*, mvt. 3, mm. 1-39, BME paradigm of the "a" small section in A large section

Rhtorical Start of Scherzo Section

Form based on BME rhetorical structure	A (Scherzo) No Displacement							B (Trio)			
	mm. 1-92	mm. 93-120	mm. 121-135	mm. 136-174	mm. 175-240	mm. 241-268	mm. 269-307	mm. 308-336	mm. 337-388	mm. 389-428	mm. 429-489
Deep Level	Beginning	Middle	Ending	Middle	Beginning	Middle	Ending	Beginning	Middle	Ending (V)	Ending (I)
Form based on themes	A (Scherzo)					B (Trio)					
	mm. 1-135			mm. 136-174	mm. 175-240	mm. 241-307		mm. 308-336	mm. 337-428		mm. 429-489
Themes	Horn call theme (a)			Waltz theme (b)	Horn call theme (a)	Langsam theme (c)		Langsam+Waltz theme (c')	Langsam theme (c)		Langsam+Waltz theme (c")
Keys	D major	B minor	D major	B♭ major	A major	G minor		D minor	A♭ major	A minor	F minor

Thematic Start of Scherzo Section

Form based on BME rhetorical structure	A' (Scherzo)					
	mm. 490-526	mm. 527-578	mm. 579-632	mm. 633-661	mm. 662-819	
Deep Level	Beginning	Middle	Beginning	Middle	Ending	
Form based on themes	A' (Scherzo)				B' (Trio)	A" (Scherzo)
	mm. 490-582		mm. 583-632	mm. 633-695	mm. 696-763	mm. 764-819
Themes	Horn call theme (a')		Horn call+Langsam theme (a")	Waltz theme (b')	Langsam theme fragment (c"")	Horn call theme fragment (a"")
Keys	D major	B minor – D major	D major - F minor – A minor	G minor – D major	D major - B♭ major	B♭ major – D major

Figure 2b: Mahler's *Fifth Symphony*, mvt. 3, Scherzo movement rhetorical and thematic structure comparison. No displacement in the initial Scherzo section

4. Langsam theme accompanied with non-beginning sign: Rhetorical displacement

When the returned horn call theme closes the Scherzo A section thematically in measure 240, an idiosyncrasy happens at the end of this large Scherzo A section in terms of missing a

clear deer-level Ending sign. In other words, it is hard to announce that the A Scherzo section is closed rhetorically due to the lack of a deep-level Ending event. So the issue is, is the following Langsam theme a continuation of the A Scherzo section (meaning part of the A section), or is it the start of the Trio? What's the relationship between the rhetorical structure and traditional formal analysis? When hearing a new theme other than the "a" section theme, it's the place to evaluate if the Trio section may begin. F horn starts a new theme from measure 241 which is different from the "a" section's horn call plus arpeggiated and scale theme, and the Waltz theme of the "b" section. The new theme set up in G minor also features a Langsam tempo (Figure 3). As mentioned above, the Langsam part introduces a new theme and new tempo, but the idiosyncrasy here is the lack of the Beginning sign and the Ending sign (a deeper-level Ending sign to close off the "a" section). The interlude material before measure 241 does not offer a satisfying Ending sign harmonically even though the melodic line offers neighboring motion as a four-measure preparation to the Langsam new theme. Even the theme itself does not come along with a clear Beginning sign with the tonic harmony (It's hard to identify the key at the beginning of the Langsam theme). The theme's melodic line played by the F horn can be identified as a G minor melody, but it only offers a relatively weaker indication of a Beginning sign melodically.

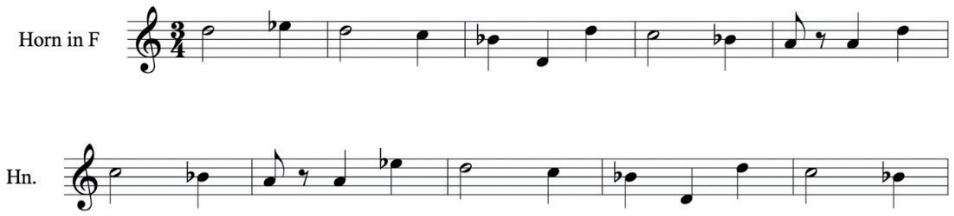


Figure 3: Mahler's *Fifth Symphony*, mvt. 3, Langsam theme at the beginning of Trio

A more significant indication is that the harmony accompanied by the melody seems to go to another function. The woodwinds (flute, oboe, and clarinet) and violin provide the arpeggiated texture and give the harmonic context by staying in the iv chord of G minor (mm. 241-246). This means the new theme starts with a predominant function rather than a tonic function. The idiosyncrasy even continues for the rest of the Langsam part. The predominant function continues in mm. 247-251 with V7/VI – VI. From measure 252, the same melodic line restates but the harmonies change to V – VI – III – vii^{o7}/V – V progression (mm. 252-259). From this progression, we can see a hint of an Ending sign due to the dominant function. The following part with repetitive texture in mm. 260-268 also establishes the sonority of vii^{o7}/V – V. By gaining energy through the repetitive texture, all instruments claim a strong sonority of the note of D and the note of F starting from measure 270 (rehearsal 10), which gives an important emphasis to the V chord of G minor.

The non-beginning feature even continues for the next 37 measures. With a slow melodic motion in mm. 277-285, the emphasis of notes D and F switches to notes of D and F# (measure 282), then ends with a PAC in D major. A slow interlude

(mm. 286-307) which is based on the Langsam theme modulates from D major to D minor at the end of the interlude (mm. 305-307). Up to this point, the entire Langsam tempo leads a section with a long Middle sign and a long Ending sign. Comparing the surface thematic statements, which indicate a traditional formal analysis of the sectional division, and a deep-level rhetorical paradigm, which provides another look at the structure according to the signs, we can find a displacement between these two dimensions, and this displacement is the main point that I am going to focus on for the entire movement's analysis. The thematic statement suggests that the Trio begins with the Langsam theme due to its new thematic features and a new key of G minor. However, the rhetorical structure in both surface and deep-level, with the signs of Middle and Ending, does not support a new Beginning of a new section. So, how should the analysis evaluate and reasonably explain this displacement? How should we understand the new interpretation of the rhetorical explanation of the form?

Since the thematic statement and the rhetorical signs create displacement when the new theme begins, seeing the music form from a new perspective can tell how Mahler generates and organizes musical materials. However, considering how to identify this Langsam section and its relationship with the previous large A section and the following Waltz theme, the interpretation is a challenge. First, the connection with the large A section is so close that it is hard to divide the Langsam from its previous section rhetorically because they share the same type and same level as the Middle sign. The same sign in the same level means the same section in my theorization. So from this perspective, the Langsam part and the previous large A section should not be separated. However, starting from the Langsam part, a new theme that is beyond the complete A section also melodically refers to the start of the Trio (a new dance). This means, as part of the two alternating dances, this Langsam combined with the upcoming Waltz, in contrast with the A section, represents the crucial feature of a Scherzo – two alternating dances.

The second challenge is that the BME paradigm shows the lack of a Beginning sign if we take the Langsam as the start of the Trio. This again leads to an interpretation that the real start of the Trio is from measure 308 (rehearsal 11), and the Langsam part functions as an introduction or bridge to the Trio. This resembles the first movement's exposition, in which the funeral march section also features a long Middle sign and an Ending sign, connecting to the primary theme. So, the paradigm of M-E-B tends to interpret the Waltz (measure 308) as the real start of the Trio. Then this explanation brings another idiosyncrasy that the Waltz theme is the "b" theme in the Scherzo A section. Thematically, or looking from the traditional formal analysis perspective, a picture of a-b-a-b structure is established, and it is hard to divide the second "b" apart.

Seeing these challenges, I interpret them as the results of two different definition criteria regarding one section from different perspectives. This is a good opportunity to analyze the form from a new perspective and unfold a deep-level picture to see the organizing logic of complex musical materials. Normally, a complete formal section's semiotic paradigm shows clues about where the section begins, where the section develops, and where the section ends through Beginning sign, Middle sign, and Ending sign. In most cases, a new theme's start accompanies a Beginning sign represented by tonic function harmonies. A special case, like this Langsam part, contains ambiguity between its thematic structure and rhetorical structure,

showing that the new theme is accompanied by a Middle sign. In other words, Mahler blurs the sectional distinguishment by using the same sign at the "background"[2] to create a global connection between the large A section and Trio. If we analyze the Langsam only based on its contrasting thematic material, its three principal groups of instruments (F horn, F trumpet, and woodwinds) and their statements demonstrate the style feature in a Classical scherzo (the Trio is the second one of two alternating dances and normally soft and employed a reduced orchestra, McKee (2005)), then the Langsam part announces the arrival of the Trio. However, a new theme without a Beginning sign indicates that at the deepest level, a developing Middle section of the movement still continues (Figure 4), which creates a displacement between thematic and rhetorical structure.

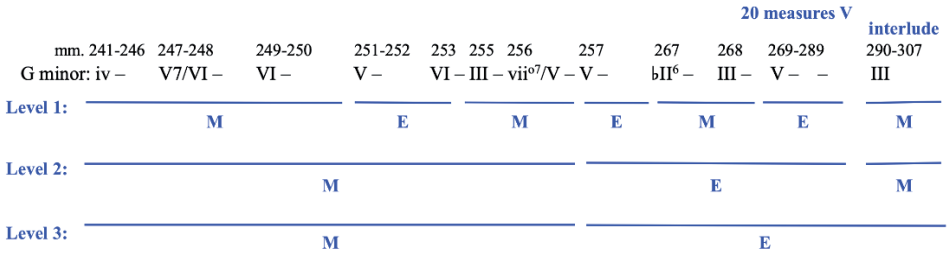


Figure 4a: Mahler's *Fifth Symphony*, mvt. 3, mm. 241-307, BME paradigm of the new theme with Langsam tempo and non-beginning sign

Rhetorical Start of Trio Section

Form based on BME rhetorical structure	A (Scherzo)							B (Trio)			
						Displacement					
	mm. 1-92	mm. 93-120	mm. 121-135	mm. 136-174	mm. 175-240	mm. 241-268	mm. 269-307	mm. 308-336	mm. 337-388	mm. 389-428	mm. 429-489
Deep Level	Beginning	Middle	Ending	Middle	Beginning	Middle	Ending	Beginning	Middle	Ending (V)	Ending (I)
Form based on themes	A (Scherzo)					B (Trio)					
	mm. 1-135			mm. 136-174	mm. 175-240	mm. 241-307		mm. 308-336	mm. 337-428		mm. 429-489
						Thematic Start of Trio Section					
Themes	Horn call theme (a)			Waltz theme (b)	Horn call theme (a)	Langsam theme (c)		Langsam+Waltz theme (c')	Langsam theme (c)		Langsam+Waltz theme (c'')
Keys	D major	B minor	D major	B♭ major	A major	G minor		D minor	A♭ major	A minor	F minor

Form based on BME rhetorical structure	A' (Scherzo)								
	mm. 490-526	mm. 527-578	mm. 579-632	mm. 633-661	mm. 662-819				
Deep Level	Beginning	Middle	Beginning	Middle	Ending				
Form based on themes	A' (Scherzo)						B' (Trio)	A'' (Scherzo)	
	mm. 490-582		mm. 583-632	mm. 633-695			mm. 696-763	mm. 764-819	
Themes	Horn call theme (a')		Horn call+Langsam theme (a'')	Waltz theme (b')			Langsam theme fragment (c''')	Horn call theme fragment (a''')	
Keys	D major	B minor – D major	D major - F minor – A minor	G minor – D major			D major - B♭ major	B♭ major – D major	

Figure 4b: Mahler's *Fifth Symphony*, mvt. 3, Scherzo movement rhetorical and thematic structure comparison, Langsam theme first appearance displacement

[2] This is a different term from Schenkerian background level. This background means not the surface, mainly indicating the harmonies and the semiotic sign represented by the harmonies.

From the above interpretation, the element of two alternating dances that is for the definition of the Trio also affects how to analyze the rest of the Trio after the initial Langsam theme. Looking through the entire long passage of mm. 308-490 until the large A section's horn call announces the return of the large A scherzo, we notice that the Langsam theme played by woodwinds or by strings' pizzicato is along with the previous "b" section Waltz theme or by itself. On a larger scale, the A Scherzo section's Horn call/Waltze and the Langsam/Waltz themes follow the AB alternation. However, considering the passage of Langsam/Waltz covers almost 200 measures, formal idiosyncrasies regarding subdivisions still raise questions. From the thematic perspective that I mentioned above, analyzing the Langsam theme encounters difficulty due to Langsam's recurrence and the combination with other themes. One possible interpretation here is that when the Langsam theme appears as pizzicato or less important than the other solo theme, then we can define this part as a dance or theme other than the Langsam. For example, when the pizzicato of the Langsam theme appears in measure 308, its dynamic and the pizzicato do not offer too much emphasis on the theme. Until measure 329, the oboe introduces the Waltz theme which comes from the "b" section of the previous large A section. This combination of Langsam theme and the "b" section's theme (mm. 308-336) offers a sense that the Trio also features an alternating format between a Langsam theme and the "b" section Waltz theme. This is the interpretation from the dance style perspective when defining subsections within the Trio. Now, the issue is how to interpret the Trio's subsections when thinking from the semiotic perspective? In other words, we have seen the recurrence of the Langsam theme by itself and by combining the Langsam and Waltz themes. Then, how do we define the start of the trio when the trio is not following a minuet structure, and how do we consider the structural weight of the different appearances of the Langsam theme? We need to conceive the explanation from the BME signs.

I have mentioned the formal ambiguity regarding the new theme accompanied by a Middle sign. From the semiotic perspective, it is crucial to notice that the D major part at the end of the large A section and the G minor part with a new theme are not interrupted by an Ending sign both on the surface and deep levels. Thus, one possible interpretation is that these two parts can be regarded as a bridge between the large A section and Trio because they share the same Middle sign. I will call these two parts features rhetorical and thematic "dualism". This means the semiotic interpretation considers the D major part and the G minor part as one part that functions as the extension of the A section's Middle sign and the introduction of the Trio. The significance of identifying the rhetorical function of this bridge is that I aim to prove Mahler's treatment of the rhetorical and thematic displacement to show an alternating feature within this Trio through the alternation between signs. Thus, the contrast that used to be reflected by the alternation between dances or themes will be created by signs (refer back to the quote at the beginning of this chapter). This method resolves the challenge that after the initial Langsam theme, the overlap of the Waltz theme and the recurrence of Langsam theme creates difficulty to identify the Trio section and its contrasting dance feature.

Up to this point, the entire Langsam tempo leads a section with a long Middle sign followed by a long Ending sign for 66 measures. Comparing the surface thematic statements and rhetorical paradigm, we can find a displacement between these two dimensions. The thematic statement suggests that the Trio begins with the Langsam theme due to its new thematic statement and a new key of G minor. However, the rhetorical structure in both surface and deep-level, with the signs of Middle and Ending, does not support a new Beginning sign of a new section with the Langsam theme.

I have identified the Middle sign as continued from the A Scherzo section to the new Langsam theme and then followed by a long Ending sign, creating a thematic/rhetorical displacement. This displacement continues to provide a new understanding of the form defined by rhetorical signs. We can also see how rhetorical signs can be used as important references to define the form when the surface-level thematic statements contain idiosyncrasies. Assuming measure 308's Langsam theme is the start of a new subsection, the evidence is the pizzicato and the key change from D major to D minor. An argument, which is also the issue when interpreting the subsection's formal design through a thematic perspective, resides in the repetition of the Langsam theme. The argument is about how to balance the roles of the Langsam theme and the Waltz theme because they overlap together starting from measure 311. In other words, the new Langsam theme, if only seen from the thematic perspective, may indicate the contrasting dance that represents the Trio section. The Waltz theme, also from the thematic perspective, is a recall of the "b" material in the Scherzo section, raising a question again of where the start of the Trio is. To resolve this argument, the BME analysis provides evidence that the Trio starts at the overlap of two themes in measure 308 due to the signs underneath. By analyzing the D minor part's harmonies in mm. 308-336, we can find that the progression only features the tonic function and (inverted) dominant function as a tonic extension in D minor, and the tonic function in F minor. Except for the last V43 – i in F minor providing a weak ending sense, the majority of this part in mm. 308-336 maintains a starting feature that indicates a deeper-level Beginning sign. This forms a contrast in terms of the paradigmatic deeper-level Middle sign of the previous Langsam part (Figure 5).

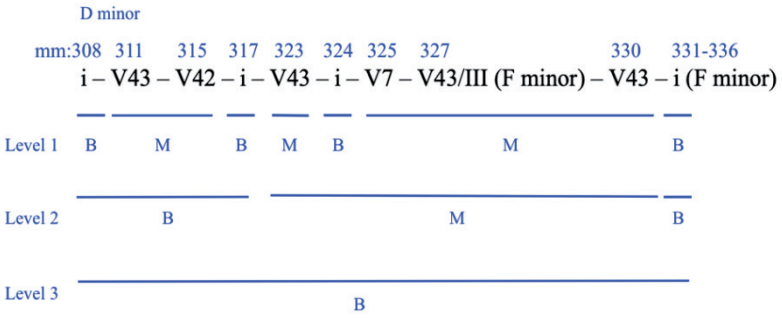


Figure 5: Mahler's *Fifth Symphony*, mvt. 3, mm. 308-336, BME paradigm of Waltz with "b" material

The Beginning sign continues with a local Ending V-I motion in mm. 337-351 due to the root position V. This part changes the key to A♭ major and continues repeating the Langsam theme. Mm. 337-351 feature V and V7 chords in root position to create a dominant area, anticipating an Ending motion V – I for a close along with the dominant pedal of E♭. However, this local incomplete Ending sign (the V chord does not resolve to I chord at the end of measure 351) does not announce a real cadential arrival to the goal of A♭ major. In measure 352, the first measure of rehearsal 12, the harmony switches to the ii predominant chord rather than the tonic. This unusual "resolution" indicates that the real function of the ii chord is not to resolve the previous V succession. In other words, mm. 337-351's dominant chords do not function as a structural Ending sign but as a continuation of the previous Beginning sign. Further, successive predominant chords after measure 351 bring back the Middle sign. Thus, the passage in mm. 337-351 connects a Beginning sign and a Middle sign. Up to this point, we can find that the Middle sign, along with the Langsam theme, continues playing, even though a Beginning sign and Ending sign are inserted in the process. It is also important to notice that the Langsam theme, starting from its initial appearance, is accompanied by the Middle sign and the Ending sign individually and is accompanied by the Waltz theme, like dual theme overlap and the Beginning sign. From a large-scale sense, the Scherzo large A section repeats the smaller "a" section to maintain the Beginning sign, then the next section uses a combination of the Langsam theme and the Middle sign continuously to hold the large section's flow. On an even larger scale, the Scherzo large A section initiates the movement with a deeper-level Beginning sign, indicating the first section of the movement. The issue is that facing the thematic/rhetorical displacement, how do we define the start of the Trio section and how do we define the structure within the Trio?

The Trio section, the second large section of an ABA distribution of the movement, stays in the deeper-level Middle sign, connecting the "bridge" of the initial Langsam theme entrance in measure 241. In other words, according to the rhetorical displacement, I define that the start of the contrasting Trio section starts from measure 308 where the Langsam theme and Waltz themes overlap. To approve this definition, my evidence is based on two perspectives. One is that between the Scherzo theme and Langsam theme, where the Scherzo A section is approaching the end, there is no clear Ending sign, neither surface level nor deep level. To distinguish sections, even though we hear musical materials change, a continuous harmonic flow means no ending, and a new section is hard to identify. Before measure 241 where the Langsam theme is introduced, we cannot find an Ending sign. So, it is hard to say the Langsam theme initiates a new section. Second, a new section's start should be accompanied by a Beginning sign, at least at the surface level. The second appearance of the Langsam theme in measure 308 starts with a Beginning sign. Here we can find how Mahler balances the relationship between thematic material and rhetorical signs. To create a connection, Mahler repeats the Langsam theme, but the overlap with the Waltz theme indicates the new section both rhetorically and thematically. In other words, the Waltz theme plays a more primary role in terms of thematic indication

from this point to show a Beginning sign with harmonies. Although Waltz's thematic material and gesture create an alternation with the Langsam theme, the continuous Middle sign on a deeper level synthesizes and tightens all material within the section of the Trio. Now, a long-standing, deeper-level Middle sign has been established, creating an expectation that it will continue along with the reappearance of the Langsam theme (Figure 6).

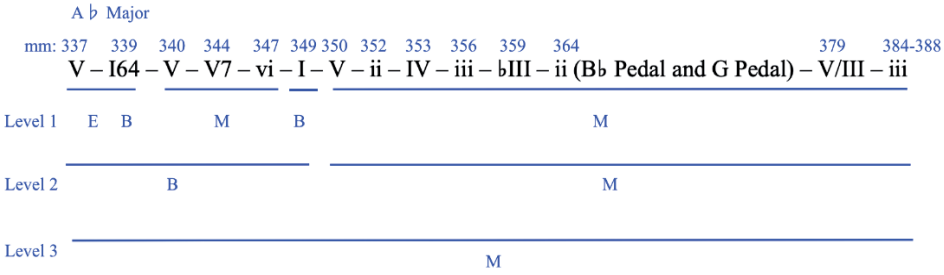


Figure 6a: Mahler's *Fifth Symphony*, mvt. 3, mm. 337-388, BME paradigm of Langsam theme

Rhetorical Start of Trio Section

Form based on BME rhetorical structure	A (Scherzo)							B (Trio)			
	mm. 1-92	mm. 93-120	mm. 121-135	mm. 136-174	mm. 175-240	mm. 241-268	mm. 269-307	mm. 308-336	mm. 337-388	mm. 389-428	mm. 429-489
Deep Level	Beginning	Middle	Ending	Middle	Beginning	Middle	Ending	Beginning	Middle	Ending (V)	Ending (I)
Form based on themes	A (Scherzo)					B (Trio)		Displacement			
	mm. 1-135			mm. 136-174	mm. 175-240	mm. 241-307		mm. 308-336	mm. 337-428		mm. 429-489
Themes	Horn call theme (a)			Waltz theme (b)	Horn call theme (a)	Langsam theme (c)		Langsam+Waltz theme (c')	Langsam theme (c)		Langsam+Waltz theme (c'')
Keys	D major	B minor	D major	B♭ major	A major	G minor		D minor	A♭ major	A minor	F minor

Thematic Start of Trio Reappear

Form based on BME rhetorical structure	A' (Scherzo)					
	mm. 490-526	mm. 527-578	mm. 579-632	mm. 633-661	mm. 662-819	
Deep Level	Beginning	Middle	Beginning	Middle	Ending	
Form based on themes	A' (Scherzo)				B' (Trio)	A'' (Scherzo)
	mm. 490-582		mm. 583-632	mm. 633-695	mm. 696-763	mm. 764-819
Themes	Horn call theme (a')		Horn call+Langsam theme (a'')	Waltz theme (b')	Langsam theme fragment (c''')	Horn call theme fragment (a''')
Keys	D major	B minor – D major	D major - F minor – A minor	G minor – D major	D major - B♭ major	B♭ major – D major

Figure 6b: Mahler's *Fifth Symphony*, mvt. 3, Scherzo movement rhetorical and thematic structure comparison, Langsam theme reappearance displacement

When the dominant 7th chord in D major overlaps the D tonic pedal from measure 696, the Langsam theme reappears as a preparation for the new section, thematically should be the second Trio. However, the thematic/rhetorical displacement again tells a

different story. I consider this starting point of the Langsam theme, not the point of the start of the second Trio because the Ending event has not been confirmed by the very last tonic harmony. From a broader perspective, this combination of the Langsam theme and the Ending sign is an important event for the process of closing off the entire movement. The reason why I treat the Ending sign in mm. 696-715 as a crucial rhetorical sign is that after this Ending sign, the harmonic change is not frequent (Figure 7). With the lack of harmonic change, the texture after measure 715 uses part of the previous Ending sign's feature – the pedal point. This means that not only does the harmony become more stable after the ending sign, but the texture "prolongs" the ending sign in terms of holding pedal points. Furthermore, the Ending sign in mm. 696-715 can be regarded as an even deeper-level Ending sign that announces the Ending event of the entire movement. In other words, although the Trio's thematic material restates and the formal structure until here can be interpreted as ABA'B', the end of the A' influences the following returning B' (or returning Langsam theme) and even the final restatement of the horn call "a" material. This influence enables the rhetorical sign of the returning Trio themes and the last statement of A themes to be a prolongation or continuation of the large and long Ending sign in mm. 696-819. Thus, although the appearance of the thematic material suggests an ABABA structure, the deepest level of the BME paradigm for the entire movement indicates the 18th-century Scherzo's three-part ABA structure (Figure 8).

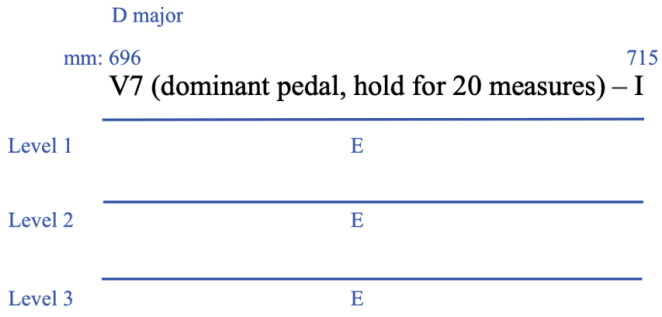


Figure 7: Mahler's *Fifth Symphony*, mvt. 3, mm. 696-715, BME paradigm of the Langsam Trio material with a long Ending sign

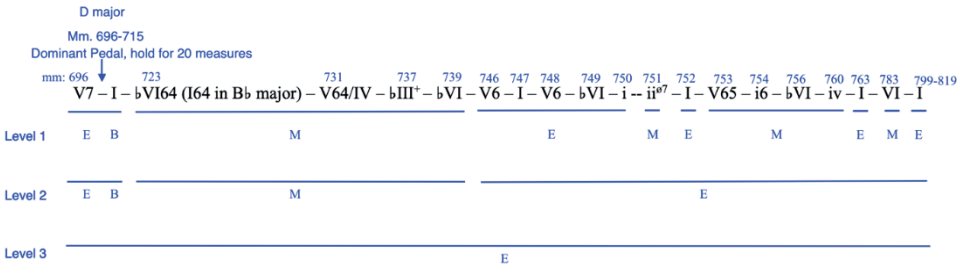


Figure 8a: Mahler's *Fifth Symphony*, mvt. 3, mm. 696-819, BME paradigm of the crucial Ending event covering the return of the Trio and the second return of A

Form based on BME rhetorical structure	A (Scherzo)							B (Trio)			
	mm. 1-92	mm. 93-120	mm. 121-135	mm. 136-174	mm. 175-240	mm. 241-268	mm. 269-307	mm. 308-336	mm. 337-388	mm. 389-428	mm. 429-489
Deep Level	Beginning	Middle	Ending	Middle	Beginning	Middle	Ending	Beginning	Middle	Ending (V)	Ending (I)
Form based on themes	A (Scherzo)					B (Trio)					
	mm. 1-135			mm. 136-174	mm. 175-240	mm. 241-307		mm. 308-336	mm. 337-428		mm. 429-489
Themes	Horn call theme (a)			Waltz theme (b)	Horn call theme (a)	Langsam theme (c)		Langsam+Waltz theme (c')	Langsam theme (c)		Langsam+Waltz theme (c'')
Keys	D major	B minor	D major	B♭ major	A major	G minor		D minor	A♭ major	A minor	F minor

Rhetorical Ending of Scherzo A'

Form based on BME rhetorical structure	A' (Scherzo)						
	mm. 490-526	mm. 527-578	mm. 579-632	mm. 633-661	mm. 662-819		
Deep Level	Beginning	Middle	Beginning	Middle	Ending		
Form based on themes	A' (Scherzo)			Displacement	B' (Trio)	A'' (Scherzo)	
	mm. 490-582		mm. 583-632	mm. 633-695	mm. 696-763	mm. 764-819	
Themes	Horn call theme (a')		Horn call+Langsam theme (a'')	Waltz theme (b')	Langsam theme fragment (c''')	Horn call theme fragment (a''')	
Keys	D major	B minor – D major	D major - F minor – A minor	G minor – D major	D major - B♭ major	B♭ major – D major	

Thematic Start of Trio 2nd Appearance

Figure 8b: Mahler's *Fifth Symphony*, mvt. 3, Scherzo movement rhetorical and thematic structure comparison, Langsam theme fragment in Scherzo section displacement

5. Conclusion

A special case, like this Langsam part, contains ambiguity between its thematic structure and rhetorical structure, showing that the new theme is accompanied by a non-Beginning sign. In other words, Mahler blurs the sectional distinguishment by using the same sign at the "background"[3] to create a global connection between the large A section and Trio. The BME rhetorical formal analysis offers a new perspective on the organization of musical materials. Under the realm of Scherzo dance genre, we can see how Mahler keeps the music flow and ABA Classical Scherzo's form indicated by rhetorical structure. Without paradigmatic and hierarchical thinking, traditional formal analyses take each movement as a container with complex musical phenomena that require adjustments to the traditional formal analyses. My perspective, however, looks inside the formal structure from a semiotic paradigm, redefining the function of certain areas according to their signs. In particular, I hear and interpret the music according to musical events, which enlarges musical moments into long passages. Through the process of this redefinition, the classical prototype hidden in the modern treatment of the musical material emerges and forms displacement with thematic statements. The

[3] This is a different term from Schenkerian background level. This background means not the surface, mainly indicate the harmonies and the semiotic sign represented by the harmonies.

paradigmatic introversive semiotic approach thus directly reveals the Classical formal prototype's capacity to embrace musical thoughts.

References

Adorno, T.W. (1991). *Mahler: A Musical Physiognomy.* Translated by Edmund Jephcott. Chicago: The University of Chicago Press.

Agawu, K. (1991). *Playing with Signs: A Semiotic Interpretation of Classic Music.* Princeton: Princeton University Press.

Agawu, K. (2009). *Music as Discourse: Semiotic Adventures in Romantic Music.* New York: Oxford University Press.

Barry, B.R. (1993). The Hidden Program in Mahler's Fifth Symphony. *The Musical Quarterly*, *77*(1), 47–66.

Brown, A.P. (2003). *The Symphonic Repertoire Volume IV: The Second Golden Age of the Viennese Symphony: Brahms, Bruckner, Dvorak, Mahler, and Selected Contemporaries.* Bloomington: Indiana University Press.

Caplin, W.E. (1998). *Classical Form: A Theory of Formal Function for the Instrumental Music of Haydn, Mozart, and Beethoven.* New York: Oxford University Press.

Caplin, W.E. (2004). The Classical Cadence: Conceptions and Misconceptions. *Journal of the American Musicological Society*, *57*, 51–117.

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (2002). *Syntactic Structures*. (2nd ed., with an introduction by David W. Lightfoot). Berlin: Mouton de Gruyter.

Dahlhaus, C. (1989). *Nineteenth-Century Music.* Translated by J. Bradford Robinson. Berkeley: University of California Press.

Dougherty, W.P. (2014). What is a Musical Sign? *Interdisciplinary Studies in Musicology*, *14*, 62–83.

Hanslick, E. (1986). *On the Musically Beautiful.* Translated by G. Payzant. Indianapolis: Hackett Publishing Company.

Hepokoski, J., & Darcy, W. (2006). *Elements of Sonata Theory: Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata.* New York: Oxford University Press.

Micznik, V. (1994). Mahler and "The Power of Genre". *The Journal of Musicology*, *12*(2), 117–151.

Monahan, S. (2011). Success and Failure in Mahler's Sonata Recapitulations. *Music Theory Spectrum*, *33*(1), 37–58.

Monahan, S. (2015). *Mahler's Symphonic Sonatas.* New York: Oxford University Press.

Nattiez, J.J. (1990). *Music and Discourse: Toward a Semiology of Music.* Translated by C. Abbate. Princeton: Princeton University.

Ninio, A. (2006). *Language and the Learning Curve: A New Theory of Syntactic Development*. Oxford: Oxford University Press.

Péteri, L. (2009). Form, Meaning, and Genre in the Scherzo of Mahler's Second Symphony.

Studia Musicologica, *50*(3/4), 221−299.

Rothfarb, L., & Landerer, C. (2018). *Eduard Hanslick's On the Musically Beautiful: A New Translation.* New York: Oxford University Press.

Samuels, R. (1995). *Mahler's Sixth Symphony: A Study in Musical Semiotics.* Cambridge: Cambridge University Press.

Schmalfeldt, J. (2011). *In the Process of Becoming: Analytic and Philosophical Perspectives on Form in Early Nineteenth-Century Music*. New York: Oxford University Press.

Temperley, D. (2003). End-Accented Phrases: An Analytical Exploration. *Journal of Music Theory*, *47*, 125−154

Walter, B. (1956). *Gustav Mahler.* Vienna: Herbert Reichner Verlag.

TEMATSKA DVOSMISLENOST I RETORIČKO POMERANJE U *MALEROVOJ PETOJ*: INTROVERZIVNA SEMIOTIČKA ANALIZA FORMALNE FUNKCIJE TEME LANGSAM U STAVU SKERCA

Apstrakt

U stavu skerca *Pete simfonije*, Maler primenjuje inovativne kompozicione tehnike koje naizgled odstupaju od klasične tradicije. Međutim, posmatrane iz semiotičke perspektive, ove varijacije mogu se sagledati kao savremena proširenja klasičnih normi unutar formalne strukture. Koristeći introverzivni semiotički pristup Kofija Agavua (2009), koji primenjuje jezičke modele u muzičkoj analizi, ovaj rad istražuje kako forma stava skerca poštuje klasičnu retoričku paradigmu Početak–sredina–kraj (PSK). Tumačenje svakog formalnog odeljka u okviru paradigme PSK omogućava uvid u funkcionalne i hijerarhijske odnose između implicitnih, labavo povezanih događaja koji ipak čuvaju logiku klasične trodelne ABA forme tipične za skerco.

U središtu ovog stava nalazi se tematska dvosmislenost u vezi sa ponovnom pojavom teme Langsam, koja stoji u kontrastu sa temama skerca i dovodi u pitanje konvencionalnu raspodelu odeljaka. Tradicionalno analizirana kao početak novog Trio-odeljka, početna pojava teme Langsam praćena je retoričkim znacima koji ne ukazuju na novi odeljak. Umesto toga, dugi znak Sredine praćen dugim znakom Kraja sugeriše tematski kontinuitet, a ne podelu. Ovo tematsko, nasuprot retoričkom, pomeranje dovodi do reinterpretacije tradicionalne formalne analize kao pojednostavljene trodelne forme. Paradigma PSK odražava pristup „odozdo naviše", dopuštajući da se tradicionalna trodelna forma izdvoji iz same muzike, umesto da se muzika nasilno uklopi u unapred zadati obrazac. Ovakvo semiotičko tumačenje klasične paradigme PSK stoga ima sposobnost da objasni složene savremene muzičke ideje kako u okviru ovog stava, tako i u čitavoj *Petoj simfoniji*.

Ključne reči: *Peta simfonija* Gustava Malera, forma, introverzivna semioza, paradigma Početak–sredina–kraj, pomeranje

THE SASE JOURNAL
ISSN 3042-2930, UDK 821.111

№ 2, 2026
pp. 111-115

Book review
UDC 371:811.111'276.6(049.32)
ORCID: 0000-0002-2997-8438

CORE CONCEPTS IN ENGLISH FOR SPECIFIC PURPOSES
Helen Basturkmen

Milica Kočović Pajević[1]
State University of Novi Pazar,
Department of Philological Sciences

English for Specific Purposes (ESP), although a branch of English language teaching (ELT), has been widely recognized in recent years as an important (separate) field of teaching and research. Its practical aspect is one of the reasons why it has become acknowledged, but theoretical contributions to the field are scarce. In recent decades ESP has expanded to encompass fields such as medicine, aviation, hospitality, law, business and many academic disciplines, while research has shifted beyond classroom practice and isolated linguistic description to address teacher training, materials design, and the function of English in multilingual contexts. There has been less recent literature on the key concepts and foundational, building block of ESP, hence this work by Helen Basturkmen has emerged. In her concise Elements volume *Core Concepts in English for Specific Purposes*, Helen Basturkmen seeks to clarify and correct taken-for-granted assumptions in the field: rather than presenting a handbook of new methods or a survey of some kind, the book concentrates on two foundational notions and two core concepts in ESP: needs analysis and specialized English and subjects them to rigorous conceptual scrutiny. Basturkmen's aims are explicit and clear: to map these concepts, expose difficulties in their practical application, and point to avenues for further research. As the author herself points out, in this book the aim is to demonstrate that ESP is not just "a practical teaching area that has evolved as an off-shoot or add-on of ELT but a distinctive field of teaching with a distinctive theoretical basis" (Basturkmen, 2025: 3).

Since the Element is deliberately concise, Basturkmen targets an audience of ESP researchers, teacher-educators, and graduate students in applied linguistics who benefit from conceptual syntheses that bridge research and practice. She places ESP historically (tracing its origins and growth since the mid-twentieth century, from the 1960s onwards), summarizes drivers of its expansion (work mobility,

[1] Email address: mkocovicpajevic@np.ac.rs
Corresponding address: Državni univerzitet u Novom Pazaru, Vuka Karadžića bb, 36300 Novi Pazar

internationalization of higher education and adoption of EMI, dominance of English as a lingua franca, policy demands), and then devotes compact, evidence-based and research-informed chapters to the two named core concepts before concluding with recommendations for future research. The volume is structured into four numbered sections: Introduction, Needs Analysis, Specialized English and Concluding Comments, each containing short overviews, discussion of practice, potential issues, and discussion questions for readers.

In the introductory chapter, Basturkmen presents the situational position of ESP, by defining ESP as teaching directed at learners' work- or study-related language requirements and by distinguishing ESP from General English teaching, which typically aims for broad proficiency. This chapter is divided into several subchapters: Overview, where the author answers the question of what ESP really is, Contexts of ESP teaching, Drivers of ESP and concludes with discussion questions. She outlines the multiplicity of ESP contexts (from pre-professional university courses to in-service workplace programmes), and argues that despite contextual variety, ESP rests on a relatively small set of core assumptions, most notably that instruction should be informed by analyses of learners' needs and that it should target the forms and practices of specialized English. The introduction also sets the methodological tone: the Element is not "a methodological cookbook" but a conceptual intervention intended to provoke reflective practice and focused research. In this part, as in all other sections of the book, the author gives some potential issues, and central to this discussion is an exploration of possible problems, particularly the fact that many ESP teachers may have only limited knowledge of the specialized language they are required to teach.

Section 2 examines the role of needs analysis (NA) in ESP, by pointing out how needs analysis and ESP are *inextricably intertwined* (Basturkmen, 2025: 5). Needs analysis has always been recognized as crucial for ESP and at the core of it (Basturkmen, 2025: 12) because it identifies learners' target communicative situations and specific language requirements, thereby directly informing syllabus design, materials development, and assessment (Hutchinson & Waters, 1987; Munby, 1978). This chapter offers a clear map of needs analysis (NA) approaches: target vs present situation analysis, task-based instruments, stakeholder consultation, and mixed-methods designs, and describes how these approaches have been used to derive course content, materials and assessment criteria. Basturkmen goes beyond procedural description to highlight the rhetorical and political dimensions of NA: choices about whose needs count, how stakeholders are consulted, and the limits of access to authentic workplace data all shape what gets taught. Although the author recognizes that NA is not exclusive to ESP and any language course can (and should) be based on NA, in ESP a set of common needs can be identified to a greater extent, compared to General English (Basturkmen, 2025: 12). In this part, the author explains the distinction between target situation analysis and present situation analysis, identifying them as two sub-analyses whose function is an extension of the NA function, that is a "gap analysis" (Brown, 2016). Although immensely important in ESP, NA also has some potential issues that the author addresses.

Namely, Dudley-Evans and St John (1998: 10) contend that the narrowly targeted nature of ESP courses makes them more motivating than General English, hence the efficiency of needs-based instruction: by concentrating on the specific language and skills learners actually require, teaching becomes more economical, which boosts learners' motivation and thereby enhances learning outcomes. In fact, motivation has been suggested as the only directly educational factor offered to explain the success of specific-purposes programs. However, as Basturkmen observes, although theory suggests ESP should be more motivating and effective than general English, there is little solid empirical research to confirm this, which is the main issue regarding NA. Additionally, the author points out practical constraints (time, institutional resources, teacher subject-knowledge) and invites readers to consider learner engagement as an important complement to the motivational rationales that typically justify needs-based ESP classes.

The main topic in Section 3 is Specialized English, a second core concept, also referred to as workplace English, i.e. language associated with a specific profession. In this section, the author emphasizes that ESP learners, who aim to enter or advance within a particular professional or academic field, must acquire the domain-specific English required in that context; consequently, ESP instruction targets specialised rather than general language competence. ESP linguistic research is primarily driven by two linked objectives: (1) to identify the characteristic linguistic forms and patterns of a given domain, and (2) to relate those forms to the communicative functions and meanings they typically realize in practice. Here, Basturkmen once again underlines vocabulary as one of the most obvious components of specialized English, by providing examples from different scientific fields. Specialized uses of English mirror the distinct values and practices of different professional and academic fields. In Section 3, Basturkmen summarizes research on specialised English (lexis, genres, discourse, pragmatics, multimodality) and links these findings to pedagogical choices. She presents a compact Framework of Linguistic Targets that helps translate corpus and genre analyses into teachable objectives (for vocabulary, grammatical choices, genre moves, and interactional routines). Potential issues that are presented within this section include limited knowledge of ESP teachers regarding the scientific field they teach in (student of law, medicine, economics), as well as materials for highly specialized areas (one of the examples provided in the book is English for dietitians). Some potential solutions are given at the end of this chapter, including but not limited to co-teaching (ESP teacher and a domain specialist), linking ESP course to a disciplinary course or simply collaboration with experts in the field to check the materials (when creating in-house materials) that ESP teachers would use in their classes. Another problematic aspect of teaching ESP that has been brought up in this section is the level of students or whether the students should have some knowledge of *Basic English* prior to studying ESP.

Although theoretical justification is limited and there are examples of ESP for elementary learners, many still claim that students should attain general English basics before beginning ESP. Indeed, Dudley-Evans and St John (1998) famously note that ESP is typically aimed at intermediate or advanced learners, and most ESP case studies report work with learners at those proficiency levels.

The conclusion restates the book's principal argument: that needs analysis and specialized English are core concepts that have not been examined sufficiently and proposes several concrete research directions. Basturkmen highlights teacher knowledge development (how instructors acquire domain expertise), low-resource approaches to needs analysis, and methods for measuring learner engagement. Contributions to the field by key scholars (such as Swales, Hyland, Gardner) are also acknowledged in the final chapter.

Basturkmen highlights the reciprocal relationship between language-focused research and ESP teaching: practical questions about the language needs of particular occupations or study programmes often drive linguistic investigations, and the results of those inquiries are routinely presented as having direct classroom applications. Some ideas about future research and solutions to potential problems are also presented in this concluding chapter. Basturkmen suggests this teacher-led repurposing as a fruitful topic for empirical study: researchers might, for example, interview ESP teachers about samples of materials they have produced, asking which academic or professional sources teachers consulted, how they interpreted and transformed that evidence for classroom use, whether they found the research easy to apply, and how their own teaching knowledge informed the reworking process. Such studies would illuminate the interface between linguistic research and instructional practice.

Her closing remarks emphasize the practical aim of this work: to encourage research that is both theoretically robust and directly applicable to teaching.

Basturkmen's *Core Concepts in English for Specific Purpose* stands out for being clear, concise and focused, successfully making foundational ESP ideas explicit without unnecessary complications. Practical features such as discussion questions and the Framework of Linguistic Targets increase its usefulness for both classroom discussion and curriculum design. The book might also have gone further in offering worked examples of low-cost NA designs or modular teacher-training activities. Its main originality is conceptual: by treating needs analysis and specialized English as objects of theoretical inquiry, Basturkmen opens productive research avenues: teacher knowledge development, engagement metrics, and pragmatic NA for low-resource contexts, that promise to connect scholarship more directly with classroom practice, bringing theory and practice closer together.

In the end, the author points out who would benefit the most from this book which is especially valuable for: graduate students in applied linguistics and TESOL who need a concise conceptual introduction; ESP/EAP researchers seeking a compact synthesis that connects methodological practice with theoretical issues; and teacher-educators looking for a short, discussion-ready text to support seminars on needs analysis, syllabus design, and teacher development. It is less appropriate as a sole resource for novice ESP teachers who require fully worked lesson plans, needs-analysis templates, or ready-made teaching materials. Overall, *Core Concepts in English for Specific Purposes* represents a valuable and timely contribution to the field, offering both clarity of thought and concise and practical insight that make it essential reading for anyone engaged in ESP teaching or research.

References

Basturkmen, H. (2025). *Core Concepts in English for Specific Purposes*. Cambridge: Cambridge University Press.

Brown, J.D. (2016). *Introducing Needs Analysis and English for Specific Purposes.* Abingdon: Routledge.

Dudley-Evans, T., & St John, M.J. (1998). *Developments in ESP: A multi-disciplinary approach.* Cambridge: Cambridge University Press.

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centred approach.* Cambridge: Cambridge University Press.

Munby, J. (1978). *Communicative syllabus design*. Cambridge: Cambridge University Press.

THE SASE JOURNAL
Publication frequency: annually
Vol. 2, 2026

Publisher
FACULTY OF PHILOSOPHY
UNIVERSITY OF NIŠ

For the Publisher
Vladimir Ž. Jovanović, PhD, Dean

Publishing Unit Coordinator
Mihailo Antović, PhD, Vice-Dean for Science and Research

Proofreading
Jelena Danilović Jeremić
Marta Veličković

Technical Editorial Office
Darko Jovanović (Cover Design)
Milan D. Ranđelović (Technical Editing)
Publishing unit (Digital Publishing)

Format
17 x 24

Print Run
20

Press
SVEN, Niš

Niš, 2026

ISSN 3042-2930.

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

821.111

THE Serbian Association for the Study of English The SASE Journal : The Serbian Association for the Study of English / editor-in-chief Jelena Danilović Jeremić. - Vol. 1 (2025)- . - Niš : Faculty of Philosophy, University of Niš, 2025- (Niš : Sven). - 24 cm
Godišnje.

ISSN 3042-2930 = The SASE Journal
COBISS.SR-ID 163096329