The Identity Thesis for Language and Music*

(version 2.1, revised January 2011)

Jonah Katz, Institut Jean Nicod (Paris) (katzlinguist@gmail.com)
David Pesetsky, (pesetsk@mit.edu)

* **Acknowledgments:** [to be supplied].

Audio: Audio for the musical examples in this paper can be found at http://web.mit.edu/linguistics/people/faculty/pesetsky/KP_paper_music_v2, and is also embedded in this PDF file. To hear the embedded audio, you must view the file in Adobe Acrobat or Reader. (The Preview program for Macintosh will not play the audio.) Click on any of the musical examples in the file.

Abstract

This paper presents and defends the following hypothesis about language and tonal music: *All formal differences between language and music are a consequence of differences in their fundamental building blocks. In all other respects, language and music are identical.*

Extending and adapting Lerdahl and Jackendoff's (1983, GTTM) model of musical structure, we argue that music, like language, contains a syntactic component in which headed structures are built by iterated, recursive, binary Merge. This is the component that GTTM called *Prolongational Reduction*, which represents hierarchical patterns of tension and relaxation in tonal harmony. We further argue that the distinct component that GTTM calls *Time Span Reduction* is a musical prosodic component (a point anticipated by GTTM itself), whose interface with the syntactic component is strikingly similar to the comparable interface in language.

Though we take GTTM as a starting point and touchstone throughout, at the heart of our proposal is a significant realignment of the GTTM model that allows us to ask several new questions. For example: is *Internal Merge* (i.e. syntactic movement) found in musical syntactic structure? We argue that Internal Merge is indeed found in music, and is exemplified by the phenomenon called the *perfect cadence.* In particular, we argue that the perfect cadence displays a clustering of properties identical to those associated with *Head Movement* in language. Building on this analysis, we argue that the output of musical syntax feeds a *Tonal-Harmonic Component* whose formal relation to musical syntax parallels that between linguistic syntax and semantic systems.

"*Etoit-il étonnant que les premiers Grammairiens soumissent leur art à la Musique, & fussent à la fois professeurs de l'un & de l'autre?*"

Rousseau, *Essai sur l'origine des langues*

1 Introduction

What is the relation between the syntax of music and the syntax of language? Although there is a rich tradition of speculative writing that seeks shared properties and common origins, the current consensus among serious researchers looks somewhat more cautious. True, there is considerable ongoing research in experimental psychology and neuroscience that concerns itself with connections and dissociations between music and language in perception and production (Patel 2003; Peretz 2006, 20-21; Peretz, to appear; Fedorenko et al. 2009, among many others). In the fields that study cognition from a structural perspective, however, the situation is quite different. Research in music theory and research in theoretical linguistics follow paths that are almost entirely independent, and there does not appear to be much effort among theoreticians to explore substantive connections between the two domains.

This state of affairs is not unreasonable, of course, if it reflects empirical reality, and in a recent survey of "Parallels and nonparallels between language and music", Jackendoff (2009) suggests that it does. In Jackendoff's opinion, although "language and music share a considerable number of general characteristics and one detailed formal one, namely metrical structure", "most of what they share does not indicate a particularly close relation that makes them distinct from other cognitive domains". He concludes by urging "caution in drawing strong connections between language and music, both in the contemporary human brain and in their evolutionary roots".

The appeal of this rather sober view is heightened by the well-known inadequacies, failures and occasionally even embarrassments of the speculative tradition mentioned above (of which Bernstein 1976 is a famous example). Lerdahl and Jackendoff (1983, henceforth *GTTM*), for example, began their book with a cautionary remark on just this point — remarks echoed recently, with only slightly more optimism, by Patel (2006):

> "Many previous applications of linguistic methodology to music have foundered because they attempt a literal translation of some aspect of linguistic theory into musical terms — for instance, by looking for musical "parts of speech," deep structures, transformations, or semantics. But pointing out superficial analogies between music and language, with or without the help of generative grammar, is an old and largely futile game. One should not approach music with any preconceptions that the substance of music theory will look at all like linguistic theory." (GTTM, 5)

> "The key to successful comparison [between linguistic and musical syntax] is to avoid the pitfall of looking for musical analogies of linguistic syntactic entities and relations, such as nouns, verbs, and the constituent structure of linguistic syntactic trees. Once this pitfall is avoided, one can recognize interesting similarities at a more abstract level, in what one might call the 'syntactic architecture' of linguistic and musical sequences." (Patel 2006, 267)

Strengthening the attraction of the sober view is the actual content of GTTM, which developed and defended a cognitive theory of the structure of tonal music grounded in the approach and outlook of contemporary generative linguistics. Even now, a quarter century later, GTTM (especially when combined with its sequel, Lerdahl 2001b; henceforth *TPS*) remains the best-developed proposal of its type, unrivaled in comprehensiveness and insight. Nonetheless, despite the inspiration that Lerdahl and Jackendoff derived from generative linguistics, they made a point of noting, in the concluding chapter of their book, that "the generative music theory developed here does not look much like generative linguistics" (GTTM*,* 307). Given the enormity of GTTM's achievement, the thoroughness with which its authors develop their proposals (Lerdahl 2009), and their explicit commitment to a perspective inspired by linguistics, the fact that the resulting model "does not look much like generative linguistics" could almost be taken as an argument in itself for a cautious view of the relation between linguistic and musical structure.

Despite these considerations, we argue in this paper for a reevaluation of the consensus view — and against the claim that the fine structure of music and the fine structure of language are distinct and largely unrelated. We will suggest that a variety of differences that look substantive at first glance are actually just artifacts of differences in approach and presentation on the part of researchers across the two domains. Our first task will be to standardize the presentation of structural proposals about language and music, so that they may be compared more productively. Though some differences will remain when this task is carried out, many of these differences can be seen as simple consequences of the fact that language, unlike music, makes use of a *lexicon*. So many other details of musical and linguistic structure will turn out to be identical that the two domains might quite reasonably be viewed as products of a single cognitive system.

Our starting point, perhaps surprisingly, will be GTTM itself, which will serve as both background and touchstone throughout this paper.[1] In the next section of this paper, we will argue that — despite the fact that GTTM was developed with linguistics in mind — its authors chose scientific goals and followed analytic hunches that actually differ considerably from the goals and hunches of the work in linguistics that they cite as inspiration. If we are correct, then wherever the "look" of the model developed in GTTM and related literature diverges from the "look" of linguistic theory, we will need to ask the following question: to what extent do these differences result from real distinctions between language and music, and to what extent do they result from non-inevitable differences in how research has been pursued in the two domains?

Our first step will be an argument that the GTTM model is already more similar to generative theories of language than one might think, and more similar than GTTM's authors themselves claimed. In particular, we will show that if we examine GTTM's formal proposals concerning music from the vantage point of current linguistic theory — and vice-versa — striking parallels emerge in the formal syntax of the two systems.

In pursuing this proposal, we will not be reviving ancient quests for total identity. We will not be arguing for nouns and verbs in music, nor for dominant and subdominant chords in language. The building blocks of linguistic syntax are **lexical items** (arbitrary pairings of sound and meaning) — which have no obvious counterpart in music. The building blocks of tonal musical structure concern **pitch-class and chord quality** — which likewise have no linguistic analogues.[2] We will suggest that

---

[1]See Jackendoff (1987, ch. 11) for a concise and non-technical overview of the GTTM model.

[2] Like GTTM, we limit our discussion to the Western tonal idiom, in which individual pitch-events, even when not explicitly harmonized, nonetheless are understood in terms of their harmonic implications. It is a commonplace to assume that many other idioms, including pre-baroque tonal music, did not have this property. Though we do wonder whether this conclusion is as well-supported as normally assumed, it is obvious that research into other musical idioms conducted in the spirit of GTTM and the present paper might significantly alter the picture of music and its relation to language offered here. (See Hughes 1991, however, for a critical view of some attempts to describe non-Western musical genres in generative terms.) Even within the Western tonal tradition, we are restricting our

what language and music have in common is not their building blocks (which are different), but *what they do with them*.

In particular, we will suggest that language and music share a common syntactic component, and that the ways in which this component interacts with other components of language and music are also highly similar across the two domains. Our hope is that whatever differences we may find between the two domains can be attributed to differences in the basic building blocks, rather than to intrinsic differences in the rules that assemble them. These differences include distinctions in how the structures of language and music are processed by distinct interpretive components. Since most lexical items come with a truth-conditional meaning, a structure whose basic elements are lexical items can be interpreted propositionally. No such interpretation is available for a structure whose basic elements are pitch configurations in a tonal space (*TPS*). Likewise, since the combinatorics of tonal harmony largely concern relations of distance or tension within a tonal space, the interpretive component of music (as we shall see) concerns itself with notions like "key" or "tonic" that are irrelevant in language. Our paper thus advances and defends the following thesis:

(1) **Identity Thesis for Language and Music**
All formal differences between language and music are a consequence of differences in their fundamental building blocks (arbitrary pairings of sound and meaning in the case of language; pitch-classes and pitch-class combinations in the case of music). In all other respects, language and music are identical.

Though the GTTM model will serve as both starting point and touchstone throughout the paper, we will also take note of an important gap in GTTM's ability to explain a central phenomenon of tonal music: the role of the **cadence** in establishing a key and providing closure for a passage or piece. In the final sections of this paper, we will suggest that this gap exemplifies a specific weakness in the logical structure of GTTM as a theory — and then argue that the gap can be remedied if we realign GTTM with linguistic theory in a manner suggested by the Identity Thesis. Our resolution to the problem of cadences, if successful, will thus constitute supporting evidence for the thesis itself.

Though most of this paper is devoted to arguments that support the Identity Thesis, we believe that the thesis makes good sense as a starting assumption, regardless of what our final conclusions about it might be. We have already noted some reasons why one might be cautious or skeptical about the effort.[3] If we had no particular cause to suspect the existence of serious, substantive formal similarities between music and language, there might be little reason to reopen the discussion. Nonetheless, despite Lerdahl and Jackendoff's overall assessment of the matter, a few formal similarities between music and language are already evident in GTTM itself (as the authors themselves note). Furthermore, the quarter century that has passed since GTTM was written has been marked by significant developments in linguistics. Ideas and discoveries about the structure of language that post-date GTTM suggest new points of similarity between linguistic and musical structure that were previously unsuspected. If only for these reasons, we consider it timely to reopen the topic.

---

attention to pieces in the traditional major and minor modes, though we see no obvious obstacles to extending the discussion to music written in other closely related systems (e.g. Western modal music).

[3] On the other hand, as Jackendoff (1977) noted in his review of Bernstein (1976), even an unsystematic, flawed approach to the problem of music and language may yield profound and useful ideas (including, in Bernstein's case, the very idea that "music can be approached in the way [linguists] are accustomed to approaching language"). An idea may be invaluable as stimulus to research even when (perhaps *especially* when) it is ultimately proved untenable in its original form.

We also believe that it is simply sound research strategy to take the Identity Thesis as a starting point for investigation. Substantive similarities between music and language are most likely to be uncovered if we construct explicit models that presuppose identity and discover where such models fail. If instead we decide to investigate music and language independently, hoping that relevant similarities will somehow come to our attention, we are far more likely to miss relevant points of identity (as well as illuminating differences). Parallel, independent investigations of complex phenomena like language and music are highly unlikely to follow properly interrelated paths of their own accord — as the history of linguistic theory and music theory already shows.

Finally, the very fact that music exists poses a number of well-known puzzles for cognitive science for which the Identity Thesis might be relevant. Whatever one's views or prejudices might be concerning how and why the human species developed language, it is not hard to think of proposals that might at least merit discussion. Music is quite different in that respect, as has been widely noted — which raises the possibility that music is merely a by-product of the coexistence of various other systems more central to human cognition (Pinker 1997, 525; Hauser & McDermott 2003; Huron 2003; Fitch 2006; Patel 2008). If this is true, we expect to find music sharing many or most of its formal properties with other cognitive systems (including language). Once again, the Identity Thesis, as stated in (1), represents a natural starting point for investigation of this question (or at least the linguistic side of it).

In the next section, we set the scene for further discussion by considering first how one might align a musical theory like GTTM with linguistic theory. This discussion is a necessary first preliminary to the development of substantive proposals about the actual properties of music and language that might bear on the Identity Thesis.

2 Overview of the issues

2.1 Levels of generative description for music

Throughout this paper, we will use the word *generative* in the sense suggested by Chomsky when he introduced it as a component of the term *generative grammar*:

> "A grammar of a language purports to be a description of the ideal speaker-hearer's intrinsic competence. If the grammar is, furthermore, perfectly explicit — in other words, if it does not rely on the intelligence of the understanding reader but rather provides an explicit analysis of his contribution — we may (somewhat redundantly) call it a *generative grammar*." (Chomsky 1965, 4)

A "generative description" of some aspect of musical cognition is one that is "perfectly explicit" in the sense indicated by Chomsky. Needless to say, in the actual world in which we live, generative description remains a goal, rather than an accomplishment, since neither perfect understanding nor even perfect explicitness has been achieved in even the most thorough proposals.

Furthermore, in music, as in language, a generative *grammar* is not the only type of generative description possible. At least four types of generative description can be imagined for music (as for language), which can be arranged in a rough hierarchy of increasing complexity and generality. (G*enerative grammar* in Chomsky's sense occupies the second level of this hierarchy.) Each level of generative description seeks to answer a characteristic set of questions:

(2) **Levels of generative description for music**

type 1. *Analysis of particular pieces:*

A listener who hears a sequence of sounds S in terms of an musical idiom I assigns one or more analyses to S. An analysis that is assigned to S can be discerned by a variety of judgments that a listener can render about S within I. What general laws define the class of possible analyses within I of a given piece?

An explicit, predictive, type-1 theory of how any sequence of sounds is analyzed in terms of a given musical idiom might be called a **generative parser** for I.[4]

type 2. ***Common properties of pieces within an idiom:***

A listener capable of assigning an analysis in a musical idiom I to a sequence of sounds S may identify or not identify S as a piece admitted by I. (For example, a classical-idiom piece without a cadence in it might be identified as outside the idiom, in that it represents at best an incomplete piece.) What general laws define the class of possible pieces in I That is, what is the grammar of I?

An explicit, predictive, type-2 theory of possible pieces in an idiom I (which generalizes over correct analyses for individual pieces, distinguishing possible from impossible pieces — and explaining the difference) is a **generative grammar** for I.

type 3. *Common properties of musical idioms:*

What is the class of possible grammars for human musical systems?

An explicit, predictive, type-3 theory of possible grammars for human musical systems might be called a **Universal Grammar** for music (**UG-M).**

type 4. *Properties common to UG-M and other cognitive systems:*

Which properties of UG-M are unique to music, and which are shared with cognitive systems generally regarded as distinct? For example, does UG-M share significant properties with Universal Grammar for language (UG-L)?

An explicit, predictive, type-4 theory thus distinguishes **music-specific components of Universal Grammar for music** from other aspects of UG-M.

---

[4] A generative parser does not need to explain the mental processes that produce a parse in real time, but may limit itself to an abstract "final-state" theory of the relation between stimuli (i.e. the musical surface) and the structures assigned to it by the mind. GTTM is a final-state theory of exactly this sort, as Lerdahl and Jackendoff explicitly note in their initial declaration of research goals:

> "Instead of describing the listener's real-time mental processes, we will be concerned only with the final state of his understanding. In our view, it would be fruitless to theorize about mental processing before understanding the organization to which the processing leads. This is only a methodological choice on our part. It is a hypothesis that certain aspects of the phenomena under investigation can be cleanly separated" (GTTM, 3-4).

Likewise, it is not part of the definition of a generative grammar that it *must* ignore the role of real-time presentation of the signal, though our proposals will mirror both GTTM and mainstream generative linguistics in not taking up this challenge. (See Steedman 1996, p. 308, for some considerations urging the opposite approach; and Jackendoff 1992, chapter 7, for additional discussion.)

The Identity Thesis as stated in (1) most directly reflects type-4 concerns. Nonetheless, it is important to note that the exploration of type-4 issues is not irrelevant even to a researcher whose interests are purely musical. Finding an answer to a type-4 question might well prove to be a prerequisite to the proper exploration and resolution of issues lower on the hierarchy. More generally, the solution to *any* lower-level question may depend on the resolution of higher-level issues — just as the solution to higher level questions depends on discoveries made concerning the lower levels.

The role of "downward" connections among the levels of our hierarchy may be less obvious than the role of "upward" connections within the hierarchy. The importance of upward connections is straightforward. Discoveries at lower levels must serve as a basis for discoveries at higher levels, or the enterprise cannot get off the ground. Productive type-3 and type-4 research, for example, must rely on the achievements of type-1 and type-2 research. We cannot discuss the common properties of musical idioms or shared properties of music and language without good comprehensive theories of individual idioms — which in turn rest on (and ultimately predict) analyses of particular pieces.

Downward connections within the hierarchy are also important, however. Discoveries made in higher-level investigations can generate productive questions for lower-level investigations, and can help limit the space of plausible answers to these lower-level questions. For example, if one is led, in the course of a type-2 investigation, to the conclusion that the best account of some property X common to a variety of pieces entails that they should also have property Y, one will then keep the possibility of Y in mind when investigating particular pieces as part of a type-1 investigation. Likewise, if one is led, in the course of a cross-idiom type-3 investigation, to propose that musical idioms in general should be structured in a particular way (cf. Balzano 1982 on possible universals of scale-systems), one would do well to keep this proposal in mind when puzzling over the properties of other particular idioms in the course of a type-2 investigation.

Finally (and most important in the context of this paper), there is a similar interaction between type-4 questions and research on lower-level issues. Imagine, for example, that research into properties of particular idioms (or cross-idiom research) suggests that musical grammars and linguistic grammars have certain specific properties in common. Such a discovery would make it reasonable to ask whether other properties are also shared. Ensuing investigation would naturally seek theories that predict the differences as well as the commonalities. Fuel for this investigation, in turn, would naturally arise from its impact on lower-level questions.

In the context of the Identity Thesis, the consequence of these observations should be clear. When investigating the structure of a particular musical passage, the grammar of a given musical idiom, or a pattern of variation across idioms, we should ask not only music-internal questions, but also cross-modal ones. Can a given phenomenon visible in music be better understood in light of some previously discovered property of language? Type-4 issues are thus not merely a topic of specialized interest, but constitute a domain of inquiry that might have as much impact on music-specific questions as music-specific research does.

2.2 GTTM in the hierarchy of generative descriptions

As we have already noted, GTTM's results as characterized by Lerdahl and Jackendoff do not favor the Identity Thesis. On the other hand, if our discussion so far is correct, it is possible that certain points of actual identity between music and language were missed in the development of GTTM — precisely because the Identity Thesis did not form part of the discussion. Though it is understandable why the authors of GTTM were wary of falling into the well-known trap of "superficial analogies between music and language", the result was a conclusion that (in our opinion) was perhaps a bit *too* Bernstein-shy. As we will argue below, certain aspects of the GTTM model, when slightly reformulated, look strikingly like models that have been developed for language — in ways that Lerdahl and Jackendoff occasionally hinted at but did not explore further.

Another reason why the GTTM model did "not look much like generative linguistics" can be seen when we consider its place in the typology of generative descriptions that we sketched in the previous subsection. Much of the literature associated with the term "generative linguistics" takes as its goal the development of a *generative grammar* (our "type 2"). This is especially true of Chomsky's contributions cited in GTTM (p.5). On the other hand, although the influence of generative linguistics on GTTM is explicit and obvious throughout the work, and although GTTM contains *elements* of a generative grammar, its actual proposals approximate the goals of a generative *parser* (our "type 1") much more closely than they resemble the goals of a generative *grammar*.

The GTTM model is explicitly designed to produce a well-formed parse (or set of parses) for any sequence of sounds.[5] The main task of the model is to produce and adjudicate among possible analyses of a given sequence No sequence is unparsable by the model. No sequence is identified or judged by the model as falling within (or not falling within) a particular idiom. The model does not identify or assign differing degrees of acceptability or felicity to pieces with respect to a musical idiom. The model also does not distinguish sequences that might be interpretable by other cognitive components from sequences that are not. Instead, the job of the GTTM theory is characterized as follows:

> "Overall the system can be thought of as taking a given musical surface as input and producing the structure that the listener hears as output." (GTTM, 11)

GTTM notes that though most musical surfaces are in principle highly ambiguous, a listener assigns only one analysis (or a small number of competing analyses) to any given piece. In light of this observation, the tasks undertaken by the GTTM model are (1) the characterization of the set of possible analyses for any given musical surface and (2) the determination of the protocols by which a particular subset of these analyses is selected for any given piece. GTTM proposed that soft constraints called "preference rules" are a key component in the selection of this subset.

An analogous set of tasks can be posed for research in the domain of language: given a "linguistic surface" (a string of words, or even a random string of vocalizations), describe how a speaker of a given language parses that string (and, in cases of structural ambiguity, determine the principles that resolve the ambiguity). In fact, there is a significant body of linguistic research on exactly these types of questions, within the subdisciplines of linguistics that study speech recognition and sentence processing. A standard topic in the study of sentence processing, in fact, is ambiguity resolution in cases where competing structural analyses are allowed by the grammar of a language (cf. Clifton & Staub 2008 for a recent survey) What this work aims to develop is a *generative parser* in our sense of the term. This research thus pursues "type 1" questions about language much as GTTM pursues type 1 questions about music.

---

[5] This characterization might be slightly too strong. Conceivably a sequence in which neither isochronous beat nor pitch can be discerned by a listener would fail to receive an analysis in GTTM. (Alternatively, it might receive an infinite number of analyses.)

As it happens, however, the development of a generative parser is *not* the topic of the particular literature in "generative linguistics" with which Lerdahl and Jackendoff explicitly contrast the GTTM theory. The work cited when Lerdahl and Jackendoff contrast their proposals with "generative linguistics" did not set as its primary goal the development of a generative parser, but rather the development of *generative grammars* for particular languages, and theories of *universal grammar* that generalize over these grammars. These are type-2 and type-3 concerns.

The distinction between the type-1 concerns of GTTM and the type-2 and 3 concerns of generative linguistics is surely one important reason why the GTTM model fails to resemble the linguistic models that Lerdahl and Jackendoff cite as precedent. Because of the type-2 and type-3 questions typically asked in generative linguistics, a typical paper in this field begins by reporting and attempting to explain some set of contrasts (in acceptability or some other measure, e.g. reaction time or brain response) among minimally differing utterances — or else discusses similar sorts of contrasts among minimally differing pronunciations or semantic interpretations of an otherwise invariant set of utterances. GTTM is silent about comparable matters for music. Indeed, Lerdahl and Jackendoff explicitly defend this property of GTTM as a natural consequence of a fundamental difference in what is "important" about music and language:

> "In a linguistic grammar, perhaps the most important distinction is grammaticality: whether or not a given string of words is a sentence in the language in question. A subsidiary distinction is ambiguity: whether a given string is assigned two or more structures with different meanings. In music, on the other hand, grammaticality per se plays a far less important role, since almost any passage of music is potentially vastly ambiguous — it is much easier to construe music in a multiplicity of ways. The reason for this is that music is not tied down to specific meanings and functions, as language is. In a sense, music is pure structure, to be "played with" within certain bounds. The interesting musical issues usually concern what is the most coherent or "preferred" way to hear a passage. Musical grammar must be able to express these preferences among interpretations, a function that is largely absent from generative linguistic theory." (GTTM, 9)

This difference in basic research goals between GTTM and work in linguistic generative grammar is of crucial importance in the present context, because it eliminates any straightforward implications for the Identity Thesis of the observation that GTTM does not look like linguistic generative grammar. We *expect* GTTM to look unlike generative grammars for language, even if the Identity Thesis is correct to the last detail— precisely because the GTTM model is *not* a generative grammar for music, but was developed to answer different sorts of questions. To properly compare a theory of music to a theory of language, the two theories should be commensurate. GTTM and models of linguistic generative grammar are not commensurate, since they reflect different types of investigations and answer different questions.

To properly investigate the Identity Thesis with generative grammars of language in mind, we therefore must try to *align* them with comparable proposals for music, so that the differences we discover will not be mere artifacts of incommensurability among models. More generally, to compare the results of distinct investigations in language and music:

(a) The **goal of the investigations** should be as close to identical as possible. If GTTM is primarily an attempt to develop a generative parser (and only secondarily a generative grammar) for music, its results may be difficult to align with the results of an investigation into language whose goal is a generative grammar (and only secondarily a generative parser).

(b) **Style of presentation and "perspective"** should be comparable. One and the same theory can often be presented from disparate perspectives — creating an illusion of differences that are not real. For example, one and the same theory can be presented in a procedural or declarative mode. Differences in notational convention can also obscure similarities in the formal structure of the theory itself. Consequently, even if the grammars of language and music should turn out to be completely identical — they might appear quite different if presented from distinct perspectives.

(c) **Differing areas of knowledge and ignorance** should be detected and acknowledged. Every existing theory of language and music is incomplete. Even if the one true theory of language and the one true theory of music should prove formally identical in the end, the incomplete theories of today might look quite different, *simply because they are fragmentary in different ways*. (As we noted above, completeness and maximal explicitness remain for now a goal of research in generative grammar, not an achievement.) Such differences may also reflect differing disciplinary traditions and prejudices in the two domains that are so deep-seated that they are hardly acknowledged — another obstacle to cross-domain comparison.

Although these desiderata look difficult in principle — in the case at hand, they may be surprisingly easy to achieve. Though the GTTM model was developed as an answer to type-1 questions about music, the structural representations and grammatical subcomponents proposed in GTTM must surely be as relevant to a generative grammar as they are to generative parser (much as units of linguistic grammar such as Noun Phrases are units relevant to the parsing problem as well). We take one of GTTM's greatest achievements to be its thorough motivation, investigation and characterization of the fundamental structures and components of musical grammar. Consequently, in asking how these structures and components might be characterized within a generative grammar, we are not starting from scratch at all, but rather attempting to fit puzzle pieces together whose existence and nature has already been established in GTTM.

Our first task, then, will be a "repurposing" of some of the components of the GTTM model within the context of a generative grammar, and we will observe that the task appears rather straightforward. It will be on the basis of the results achieved in our "repurposing" that we will be able to take productive issue with GTTM's claim that type-2 considerations relevant to issues of acceptability and degrees of grammaticalness are not "important" in music, when we turn to the syntax of cadences in section 5.

3 Aligning the theories

3.1 Should we attempt the alignment task?

Although the overall model proposed in GTTM looks unlike linguistic generative grammars in some respects, there are also points of similarity. To begin with, several of GTTM's structural representations resemble comparable proposals for language. As Lerdahl and Jackendoff themselves note, for example, GTTM's hierarchical representation of *metrical structure* is formally identical to the representation of linguistic stress proposed by Liberman & Prince (1977) and developed in much subsequent work (e.g. Hayes 1995).[6] This particular point of contact will not be central to our

---

[6] The "phonetic implementation" of the representations in the two domains differs, however, since there is no obvious isochronous beat in the areas of language discussed by Liberman & Prince (though

discussion here (though its relevance to the Identity Thesis should be clear) — but other similarities will be of crucial importance, and will serve as starting points.

For example, Lerdahl and Jackendoff also point out (GTTM, ch. 12) important points of similarity between a component of their model called *Time-span Reduction* and the systems developed by Selkirk (1980a; 1980b) and others as models of *prosody* in language (how sentence-level phonology and syntax interact). As we will see, both Time-span Reduction and the linguistic representation of prosodic structure are characterized by hierarchically nested phrasal domains, each headed by a unique event at every hierarchical level. In both systems, the property of being a head involves some notion of rhythmic prominence, associated with the percept of *stress* in language, and with rhythmic and harmonic properties in music.

Another component of the GTTM model, *Prolongational Reduction*, features binary-branching hierarchical structures that are also headed ("endocentric"). These structures are in many respects formally identical to representations currently posited for linguistic syntax. In this domain, GTTM's ideas about music preceded by some years the development and exploration of similar ideas in linguistics by Kayne (1984), Larson (1988), Chomsky (1995a; 1995b, ch. 4), and others. Consequently, this important domain of similarity went largely unnoticed in GTTM.

Despite these points of similarity, many of the most basic properties of the GTTM model differ sharply from anything proposed in generative models of language — suggesting the possibility of daunting obstacles to any imaginable alignment of the two models. Our presentation of the Identity Thesis in (1) already called attention to the fact that the raw ingredients and basic building blocks of music and language are quite distinct. Music is analyzed in terms such as pitch, key, scale, interval, chord type, and cadence. Language, on the other hand, is analyzed in terms such as phoneme, morpheme, word, part of speech, and agreement. None of these basic objects and relations from each domain has an obvious counterpart in the other.

The differences between GTTM and linguistic theory do not stop there, however. The overall architecture of GTTM looks quite unlike the architecture of generative grammars for language. The GTTM model has four major components (of which we have already mentioned two): *Metrical Structure*, *Grouping*, *Time-Span Reduction*, and *Prolongational Reduction*. Although the analyses presented in GTTM are sometimes presented as unidirectional mappings from representations in one component to another, the "official" presentation of the model posits a more complex pattern of information flow among the four components. Furthermore, the interaction among hierarchical representations in the various components is characterized in both bottom-to-top and top-to-bottom terms (a point we will clarify below). Language, in contrast, is often characterized as having three major components that interact in a strictly unidirectional ("Markovian") fashion. Lexical items are assembled by the *syntax* into hierarchically structured complex larger units by the recursive operation *Merge*. At a particular point in the assembly of syntactic structures (called *Spellout*), information relevant to semantic interpretation and phonological interpretation is transferred from the syntax to the *semantics* and to the *phonology*. There is very limited information flow from the phonological and semantic components back to the narrow syntax (perhaps none; but cf. Fox 1995, 1999), and the incremental, bottom-to-top assembly of syntactic structures is strictly mirrored in the compositionality of semantics and phonology (cf. Chomsky and Halle 1968; Bresnan 1971; Marvin 2003; Michaels 2009).

---

work in progress by Keith Johnson and colleagues suggests that this might be an oversimplification; see Tilsen & Johnson 2008).

A further difference is the presence of *filtering* interactions prominent in linguistic models, but absent from the logical structure of GTTM**.** Distinct components of linguistic models may impose conflicting requirements on particular structures. Such conflicts may be detectable as a judgment of deviance, as the absence of an otherwise expected meaning, as the unavailability of an otherwise expected pronunciation, or as a behavioral or brain response observed under laboratory conditions. No such interactions are posited in the GTTM model, in keeping with Lerdahl and Jackendoff's belief that judgments of deviance ("grammaticality") do not arise in music as they do in language. Reflecting the type-1 nature of GTTM, no component of the model filters the representations of a distinct component, as happens in linguistic models. Instead, each component can make something of the representations that feed it — just as the system as a whole can make something of any sequence of sounds fed into it.

At this point, despite the points of superficial similarity between GTTM and the linguistic model sketched above, one might be persuaded by the existence of such substantial differences to abandon the program of aligning the two theories as the first step towards investigation of the Identity Thesis. Instead, one might plausibly conclude that any serious attempt to pursue the possibility of formal similarities between language and music will need to begin with a fresh start of some sort — either (1) setting aside GTTM, or (2) setting aside established linguistic theory, or (3) both. Indeed, something like the first strategy is urged by Rohrmeier (2008), for example, whose work we return to in section 6.2; something like the second is suggested (in passing) in a more recent paper by Jackendoff (2006, 9); and works such as Steedman (1996) can be understood as simultaneously advocating new foundations for both music theory and linguistic theory (for Steedman, combinatory-categorial grammar as the proper framework for both domains).

We will suggest a more conservative conclusion. Despite the apparent deep differences between GTTM and the linguistic theory outlined above, we believe that the two models can be aligned successfully, with interesting results. We argue in particular that...

1. once we realize that the GTTM model is a generative parser that can inform the development of a generative grammar — but is not itself a generative grammar; and

2. once we recast the technical details of GTTM in a perspective and approach more familiar from work in generative linguistics — and vice versa; and

3. once we identify some distinct areas of incompleteness in GTTM vs. linguistic generative grammar, and begin to remedy these gaps,

...the GTTM model and existing linguistic theory end up looking much more alike than generally believed. This conclusion in turn suggests that the future development of a common musical and linguistic model can build successfully on the achievements of both. Furthermore, where irreducible differences between the two proposals remain even after our efforts at alignment, we will be able to conclude that we have discovered *real* differences that future research will need to come to grips with — not pseudo-differences that merely reflect distinctions in research perspectives and goals.

3.2 The GTTM model

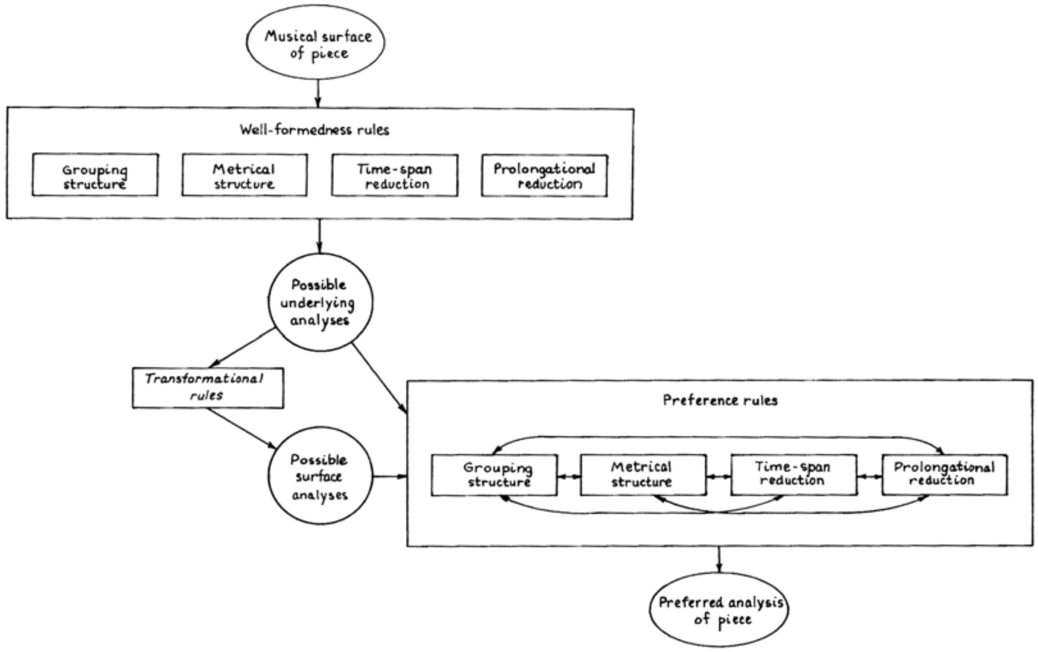
Both GTTM and current work in linguistics present their domains of inquiry as a system of *components*. A component is a cluster of rules and structural representations that appear to form a coherent unit within the overall theory and transfer information to or from specific distinct components. In this section and the next, we sketch the components of each theory.

The four components of the GTTM model each contribute to the analysis of a given musical surface in a different manner. As GTTM describes them:

1. "***Grouping structure*** expresses the hierarchical segmentation of the piece into motives, phrases, and sections.
2. "***Metrical structure*** expresses the intuition that the events of the piece are related to a regular alternation of strong and weak beats at a number of hierarchical levels.
3. "***Time-span reduction*** assigns to the pitches of the piece a hierarchy of "structural importance" with respect to their position in grouping and metrical structure.
4. "***Prolongational reduction*** assigns to the pitches a hierarchy that expresses harmonic and melodic tension and relaxation, continuity and progression." (GTTM, 8-9)

Each component produces candidate representations whose properties are characterized by a set of *well-formedness rules.* Among the class of well-formed representations for a particular piece within each component, only a subset are normally preferred by a listener — often, just a single representation. A collection of soft constraints (*preference rules*) for each component picks this subset. The preference rules of each component, as described in GTTM, make reference to the properties of the other components. In the "boxology" of the theory, the components are therefore represented as characterizing co-present representations, with information-flow proceeding in every possible direction within the 'preference rules' box:

(3) **GTTM's presentation of the model**

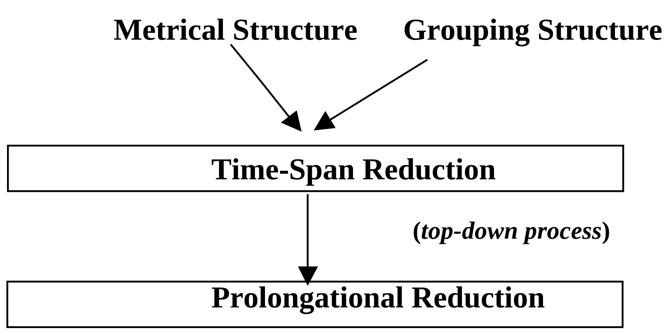


In the analysis of actual pieces, however, the relations among the components prove more asymmetric than the diagram in (3) might suggest. First, the auditory stream is parsed into groups and assigned an underlying meter in distinct **Grouping Structure** and **Metrical Structure** components. Information from these two rhythmic components together is used to establish **Time-Span Reduction (TSR) —** a hierarchy of headed domains that establishes event prominence. In general, TSR looks much like Grouping Structure, but a head is assigned to each group and constituency is determined by metrical structure below the level of the smallest Grouping Structure constituent.[7] In addition, "splicing transformations" apply when one or more event is shared by two adjacent groups (a circumstance we will not discuss in this paper).

The structure established in the TSR component is mapped into the distinct binary-branching, headed, hierarchical structure called **Prolongational Reduction (PR)**. The mapping process scans a TSR structure *top-down*, building the corresponding PR from the top down as well. As a default, PR and TSR are isomorphic, but re-bracketing is possible (yielding regions of non-isomorphy) under specific conditions.

Since the prolongational reduction is the harmonic interpretation of the piece, non-isomorphy between TSR and PR corresponds to the intuition that a melodic or rhythmic phrase boundary sometimes fails to coincide with a harmonic phrase boundary. The top-down procedure helps constrain non-isomorphy of this sort, in ways that we will discuss in detail shortly.[8]

(4) **A "practical boxology" for the GTTM model**



---

[7] Our characterization glosses over some complexity introduced by phasing mismatches between metrical and grouping constituents, which will not concern us here.

[8] GTTM also includes several constraints that would have to be described as having a 'look-ahead' property if we adopt our "practical boxology" for the GTTM model, rather than GTTM's. These are constraints that favor representations at one level constructed so as to be consistent with preferred structures at the next level. Such constraints do not seem to play a major role in any of GTTM's actual analyses, and we will not discuss them further in this paper.

3.3 A model for language

The linguistic theory we consider here[9] takes as its fundamental building blocks a set of linguistic objects stored in long-term memory, called the *Lexicon*. The Lexicon is a structured list of assemblies of features (sound, meaning, number, grammatical gender, etc.) called *lexical items*. The *syntax* takes a set of lexical items from the lexicon as input, and applies an iterating rule ***Merge*** that recursively forms two-membered sets (*constituents*), each member of which may be either (1) a lexical item or (2) a set formed by a previous instance of Merge.[10] One member of each set is designated as the constituent's *head*.

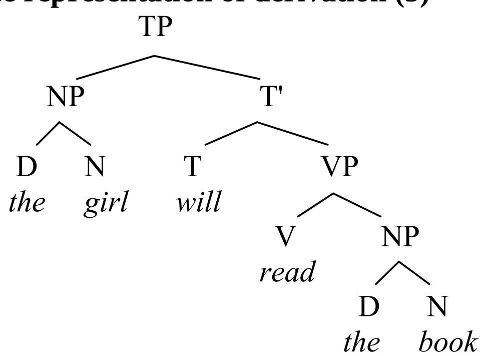
Example (5) shows how the English sentence *The girl will read the book* is assembled by iterated recursive Merge, omitting some details. The head of each constituent is indicated with an underscore. The right-hand column shows which ingredients of each instance of Merge are lexical items, and which were formed by previous instances of Merge:

(5) **Derivation by Merge of *The girl will read the book***

a. Form the set {*the*, *book*} (lex. item, lex. item)
b. Form the set {*read*, {*the*, *book*}} (lex. item, a.)
c. Form the set {*will*, {*read*, {*the book*}}} (lex. item, b.)
d. Form the set {*the*, *girl*| (lex. item, lex. item)
e. Form the set {{*the*, *girl*}, {*will,* {*read*, {*the*, *book*}}}} (d., c.)

The sequence of Merge operations sketched in (5) may also, of course, be presented as a tree like (6), in which each application of Merge in (5) corresponds to a pair of *sister nodes* whose mother is labeled in a manner that identifies one of its daughters as the head. We return shortly to other issues raised by the node labeling in (6).

(6) **Tree representation of derivation (5)**



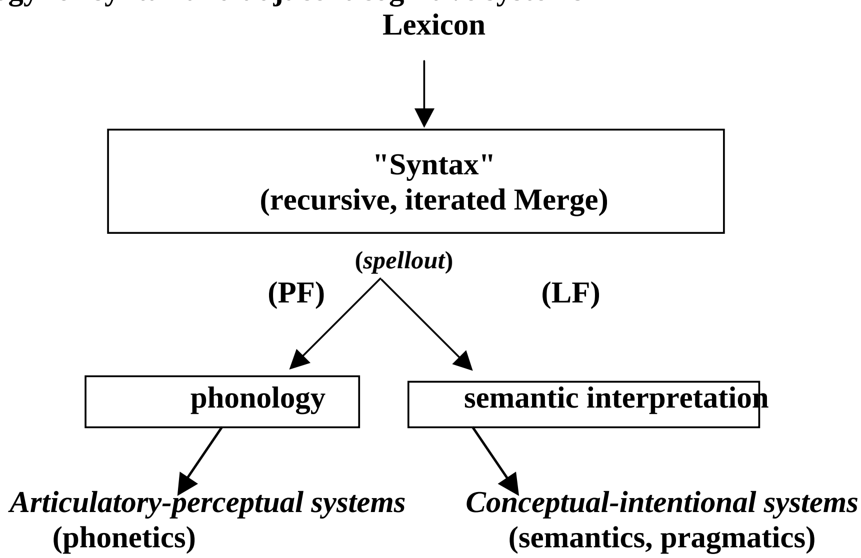
---

[9] We dwell on a number of elementary points in this section to assist non-linguist readers, and also to highlight those key properties of language whose correspondents in music will be at stake in this paper. For a brief discussion of alternative views of linguistic syntax, and their potential relevance to the issues discussed in this paper, see n. 64.

[10] We discuss so-called *Internal Merge* (= syntactic *movement*) in section 5. A more complete analysis of the sentence in (5) than we offer here would involve at least one instance of Internal Merge as well (raising *the girl* from a position within the VP).

The structure produced in the syntax by iterated Merge is next separated into two substructures. A *Logical Form (LF)* representation is formed by stripping away the phonological features of the lexical items that merged in the syntax; LF ultimately feeds *semantics* after some further computations. A *Phonetic Form (PF)* representation is formed by stripping away the semantic features of the lexical items that merged in the syntax; PF ultimately feeds phonetic systems after further phonological computations. The point in the derivation where this separation occurs, as mentioned earlier, is called ***Spell-out.***

(7) **A boxology for syntax and adjacent cognitive systems**



To successfully map to phonology or semantics, a structure produced by the syntax must meet the *well-formedness conditions* imposed by the LF and PF interfaces, respectively.[11] Failure to satisfy these conditions produces the percept of *deviance* mentioned above, which linguists indicate with an asterisk. For example, if the syntax leaves a noun phrase in a "non-case-marked position" (such as the complement to a noun or adjective), an otherwise successful syntactic structure may be blocked from mapping successfully to PF by the *Case Filter* (Vergnaud 2006, Rouveret & Vergnaud 1980, Chomsky 1980):

(8) **Effects of the Case Filter on the acceptability of noun phrases**

a. The Romans destroyed **the city**.

b. [The Romans' destruction **of the city**] was a tragedy.
c. *[The Romans' destruction **the city**] was a tragedy.

d. The earthquake was **destructive to the city**.
e. *The earthquake was destructive **the city**.

The sensation of deviance associated with (8c, e) is attributable to the *filtering effect* of the interfaces between syntax proper and adjacent systems. If the result of Merge does not place a nominal in a legal position, the adjoining systems reject or fail to interpret what the syntax has produced.

[11] Chomsky (1995b) has speculated that such conditions might be limited to those determined by intrinsic properties of the interface itself, but this point will not concern us.

Anticipating later discussion, let us focus a bit more on the box labeled "phonology" in (7). The phonological computations that apply to the PF output of the syntax include both word-level and sentence-level processes. Both types of phonology deal with both *segmental* properties of words (e.g. what distinguishes /p/ from /m/) and *supra-segmental* properties (e.g. grouping, pitch and stress), i.e. **prosody.** The prosodic structure of a sentence is derived by converting the headed, hierarchical, binary syntactic structures produced by Merge into a second kind of headed, hierarchical structure (***prosodic structure***) by a derivational process — or by a parallel (as opposed to serial) process of evaluation against a set of violable constraints (depending on theoretical approach). In the next section we compare this process to the mapping between TSR and PR in music.

4 <u>The syntax and prosody of language and music</u>

4.1 What aligns with what?

Musical TSR aligns with linguistic prosody: As Lerdahl and Jackendoff themselves discuss (GTTM, ch. 12), the TSR representations of GTTM are similar in a number of respects to representations that linguists have proposed for prosodic structure, and the interpretation of both TSR and prosodic structure invokes notions of relative prominence. Let us therefore investigate the possibility that these components should be aligned, so that TSR is the musical counterpart to linguistic prosody.

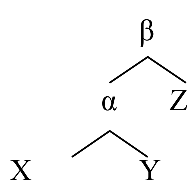
Musical PR aligns with linguistic syntax: Another obvious candidate for alignment, particularly in light of research more recent than GTTM, is the syntactic structure produced by Merge (for language) and GTTM's Prolongational Reduction (for music). The linguistic structures produced by Merge are similar to PR in two fundamental respects. First, both can be understood as forming binary-branching, headed, acyclic directed graphs. In addition, the constituents of PR and linguistic syntax are headed, and thus encode structural relations between elements that are not necessarily *string-adjacent* (i.e. adjacent in the musical or linguistic surface).

4.2 Properties of PR structures

GTTM's notation for PR representations looks somewhat different from standard syntactic conventions in linguistics, but the actual information conveyed is identical. The formal identity of PR and linguistic phrase structure is obscured by the differing graphical conventions adopted in GTTM and in the standard practice of generative syntax. In particular, though a standard "syntactician's tree" like (6) offers a convenient representation of the sequence of Merge operations in (5), it represents some information redundantly. When the redundancies are eliminated, the formal identity of the two systems becomes apparent.

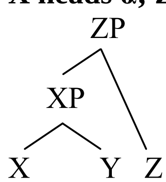
Imagine that X merges with Y, forming the set α={X, Y}, and that α and some other element Z merge next, forming the set β={α, Z}. This derivation can also be presented in the form of the tree (9), in which X and Y are daughters of α (= are *immediately dominated* by α), and α and Z are daughters of β (= are immediately dominated by β):

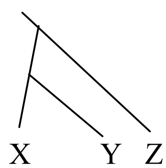
(9)



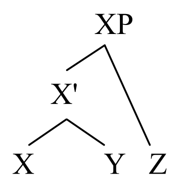
The diagram in (9) does not yet indicate the head of α or the head of β. This is where the diagrammatic conventions of GTTM differ from common practice in linguistics — but the difference is merely a matter of graphics. GTTM indicates the head of a constituent geometrically: the head H of a constituent C whose mother is M is marked by a line-segment HC that forms a 180º angle with a line-segment MC. (A line-segment from C to a non-head daughter forms some other angle with MC.) In plainer English, the line from the head continues on straight above the branch, while the line of the non-head terminates at the branch. Common practice in linguistics, on the other hand, indicates the head of a constituent C by a *label* written at C. A constituent whose head is H[12] is conventionally labeled **H'** (an intermediate projection of H) if it is dominated by another projection of H, and HP (the maximal projection of H) otherwise. The constituent structure in (9) will thus be represented with the (a) diagrams in (10)-(13) in linguist's notation, and with the (b) diagrams in GTTM notation:

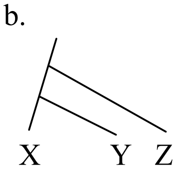
(10) **X heads α; Z heads β**

a.


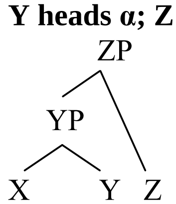
b.


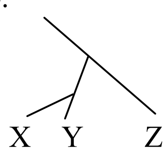
(11) **X heads α and β**

a.


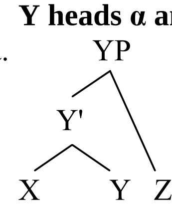
b.


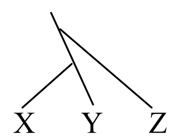
(12) **Y heads α; Z heads β**

a.


b.


(13) **Y heads α and β**

a.


b.


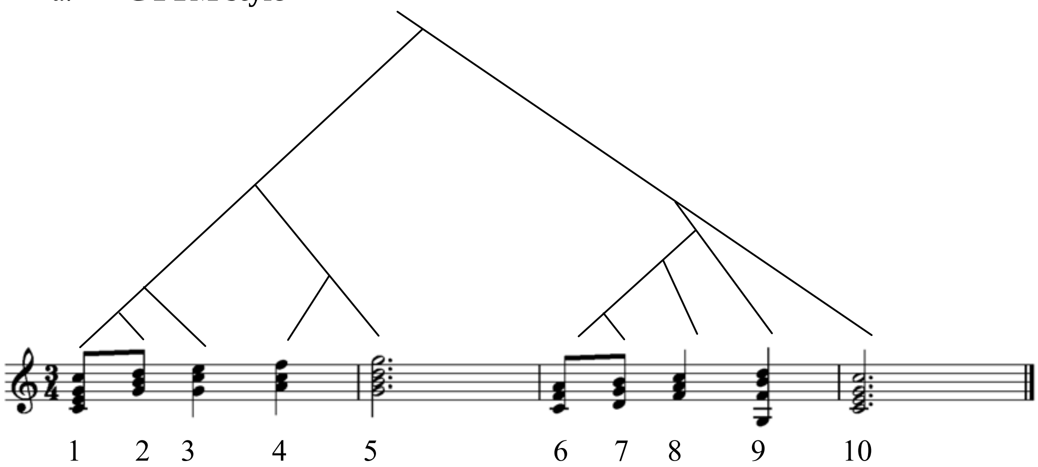
Example (14a) shows a PR structure for a toy melody in GTTM notation, while (14b) conveys the same information in linguist's format. The head of each node formed by Merge (i.e. each non-terminal node, what a linguist would call a "phrase") is identified here for convenience by numbering the chords from left to right. The question of how these nodes are actually labeled will be taken up in section 6.

---

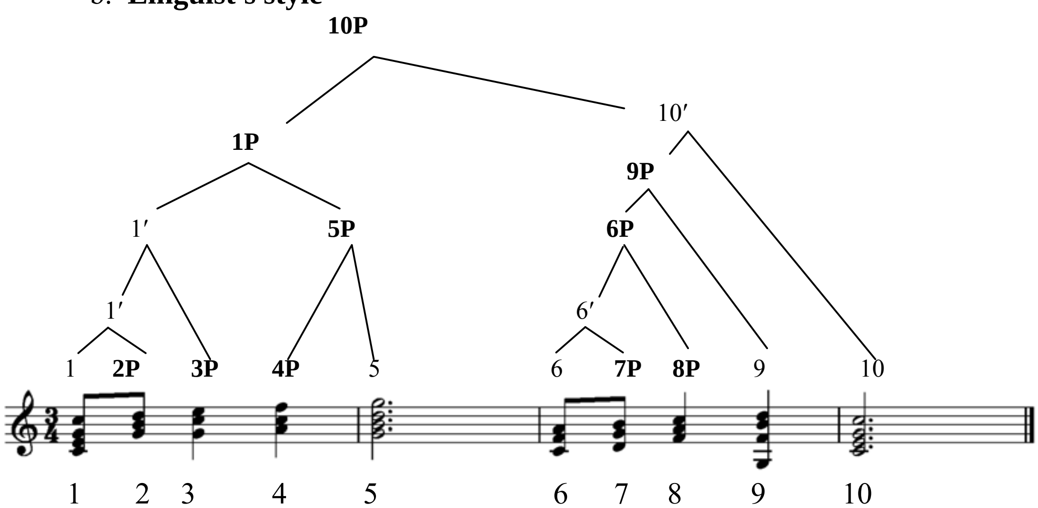
12 We will avoid the common linguist's notation **H**º for the head of a phrase, to avoid confusion with the use of the degree sign in music theory that indicates a "diminished triad".

(14) **PR structure for a toy melody**

a. **GTTM style**

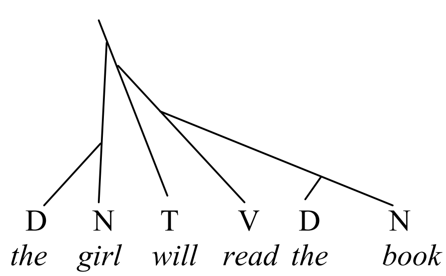


b. **Linguist's style**



Conversely, a syntactic derivation for a sentence like (5), for which the common-practice linguist's diagram is (6), can equally well be represented with a GTTM-style diagram as in (15):

(15) **GTTM-style tree corresponding to (6)**



Though the mapping between GTTM-style diagrams and common practice in linguistics is straightforward, the notational difference between GTTM and linguistic practice has some significance nonetheless, we believe. Variations in the notation with which one expresses a theory can influence one's thinking about the actual topics under investigation**.** Even when different sorts of diagrams represent exactly the same information (as is the case here), the differences among them may reflect and reinforce differing working hypotheses or hunches about the kinds of phenomena one expects to model. Differences of this sort between GTTM and common practice in linguistics arise in two important domains: the relevance of *projection level* and the amount of *information* that projects from a terminal node to the constituents that it heads.

4.3 Projection level in PR

Although one can certainly distinguish among zero-level projections, intermediate projections and maximal projection (H, H' and HP) even using GTTM notation, the explicit labeling of these nodes in standard linguistic diagrams reflects the fact that the distinctions among these three levels of projection is believed to be linguistically significant. In particular, a variety of independent phenomena of language are sensitive to the distinction between maximal and non-maximal projections, while other phenomena care about the distinction between zero-level and non-zero-level projections. For example, it is maximal and zero-level projections that may undergo the process of syntactic *movement* (also known as "Internal Merge", discussed at length in section 5) — while intermediate-level projections do not. Furthermore, the laws governing movement of maximal vs. zero-level projections appear to differ.

By contrast, the absence of such projection labels in GTTM's notation reflects a hunch rooted in Heinrich Schenker's notion of *reduction* (Forte 1959, 14ff) — that no musical principle will be sensitive to such distinctions among projection levels. All things being equal, the GTTM model predicts that the combinatory potential of a given C-major triad, for example, should be identical to the combinatory potential of every larger PR constituent headed by this triad. It is for this reason that one can generally omit the non-heads of any PR constituent in a tonal melody and the result, when performed, will more or less still "sound like music". (We return to this observation in section 4.5.) This Schenkerian hunch forms part of GTTM's *Strong Reduction Hypothesis*, in particular as extended to PR in GTTM's *Prolongational Hypothesis 1* (GTTM, 211) . This hypothesis states that all events in a piece are heard as a hierarchical segmentation of the piece into tensing and relaxing motions, with the more "relaxed" daughter of any node generally chosen as its head. Just as the same tensing and relaxing relations are present at each hierarchical level, the characteristics of specific pitch events that are relevant to the *relaxed/tense* distinction remain constant at each hierarchical level.

In PR, this property helps determine which daughter of any given node is picked as its head. Indeed, in this area of music theory, the Schenkerian hunch appears to be correct. If in a given key (for example C-major) the chord C is deemed to be more "relaxed" (e.g. closer to the tonic) than the chord G, then at any point in hierarchical structure where a node headed by C and a node headed by G are sisters, the same principles will choose C as the head of the constituent, and G as the non-head — whether we examine the very bottom of the tree (where the C and G pitch-events would be adjacent) or some higher point (where they might not be). The choice of head is thus oblivious to distinctions among zero-level, intermediate and maximal projections that loom large in linguistic syntax.

We will argue below, however, that other aspects of musical grammar do care about such projection distinctions in PR. First, as we shall shortly see in section 4.6, GTTM's own proposals concerning the mapping between TSR and PR requires the notion of **maximal projection** (though the text of GTTM itself does not call attention to the importance of this notion). In addition, we will propose that the phenomenon of *cadence* in music involves a process of syntactic *head movement* familiar from linguistics — a process that crucially distinguishes **zero-level projections** from all other projection types. If we are correct, the differing hunches that lie behind the distinct notational choices of GTTM and standard linguistics do not reflect any genuine empirical differences between the two domains. Music, just like language, might indeed be sensitive to the distinctions among maximal, intermediate and zero-level projections.

4.4 Projection of information from head to phrase in PR

Although we will argue that certain phenomena in both language and music care about differences in projection-level, there certainly are other phenomena that are indifferent to such differences, treating all projections of a given H the same way. It is these "Schenkerian" phenomena in both music and in language that help to motivate the notion of projection and headedness in the first place. This observation leaves open important questions of detail, however. Exactly *what* information about a given head H is also present at H' and HP? *How much* information projects from head to phrase?

Standard diagrammatic conventions in linguistic syntax express the conjecture (often unstated) that what projects is limited to part-of-speech information. It has always been obvious, however, that this proposal is too weak. Other information, for example a noun's number and grammatical gender, appears to project as well, as the syntax of agreement, for example, makes clear.

In contrast, the conventions adopted in GTTM for music express the Schenkerian conjecture that *all* information about the head of a phrase projects. As it happens, this very conjecture was also advanced within linguistics a decade ago (and more than a decade after GTTM) by Chomsky (1995a; 1995b, ch. 4), who proposed (under the slogan *bare phrase structure*) that the label of a projection of a given lexical item is nothing more than the lexical item itself. If the Bare Phrase Structure hypothesis is correct, *read the book* is actually a *read*-phrase rather than a Verb Phrase (VP). For language, the Bare Phrase Structure view is probably too strong. Certain properties of the heads of phrases, though relevant to other components of the grammar, do *not* project to the phrasal level in the syntax. For example, phonologically relevant information such as "ends with a fricative consonant" or "stressed on the penultimate syllable" appears to be irrelevant to syntactic phrases. Syntax is also oblivious to morphological information such as "forms its past tense by adding *-t*" or "belongs to the third conjugation". No process of syntactic agreement, for instance, cares about conjugation-class in languages like Spanish that have such phenomena — even though conjugation-class is of great relevance to the phonological and word-formation components of the linguistic grammar.

In section 6.2, we argue that similar questions about the projection of information from head to phrase arise for music, and that in music, just as in language, some properties of a head project to the phrasal level while others do not. Because the raw materials of music are different, the details of this distinction are correspondingly different. We will argue that what projects in music is pitch-class information. Other information, such as harmonic function, does not — though once again such information is highly relevant to other components of the musical grammar.

For the moment, however, we put aside the questions of projection-level and information projection that are raised by the differing notational choices of GTTM and linguistic syntax, and stress once more the formal *identity* of the two proposals. Both involve constituent structure attributable to binary, iterated Merge, with one element of each Merge operation always designated as the constituent's head.

It is in fact the formal identity of the two proposals that renders particularly puzzling one of the most striking *differences* between GTTM's PR and the syntactic component of linguistic theory: the ways in which musical PR and linguistic syntax interact with other components of music and language — as can be easily seen by comparing the charts in (4) and (7). In the next section we take up this question in detail, in the context of the proposal with which we began the current section: that not only is GTTM's PR the same system as linguistic syntax, but GTTM's TSR is also the same system as linguistic prosody (part of the "phonology" component in (7)). We will argue that the way in which PR interacts with TSR is essentially identical to current proposals concerning the mapping in language between syntactic and prosodic structure — despite the deep differences in how these interactions are presented in GTTM and in the linguistic literature. To see this, however, we will need to examine the formal properties of both proposals in some depth, and recast both of them in a way that reveals the otherwise hidden similarities.

4.5 Why both PR and TSR?

To understand the significance of the question taken up in this section, it will be useful to first review some of the musical intuitions that seem to motivate the coexistence of PR and TSR as properties of musical structure. A particularly useful discussion of this point is provided by Jackendoff (1987; chapter 11), who stresses the well-known phenomenon of musical *variation* as a good source of evidence for the coexistence of these two structures. Our discussion builds on his.

Consider once again the toy melody in (14), and the PR structure assigned to it there. Headedness in this structure corresponds to the patterns of harmonic tension and release found in this piece, which are straightforward: chords built on the first scale degree (the tonic, here C) are maximally relaxed, while chords built on other scale degrees are more tense. Independently, when chords built on the fourth and fifth scale degrees (F and G, the *subdominant* and *dominant*, respectively) are sisters, the latter is picked as the head — a general property of Western tonal music, relevant to the notion "cadence" discussed in section 5 below (GTTM 205, 211 ff).[13]

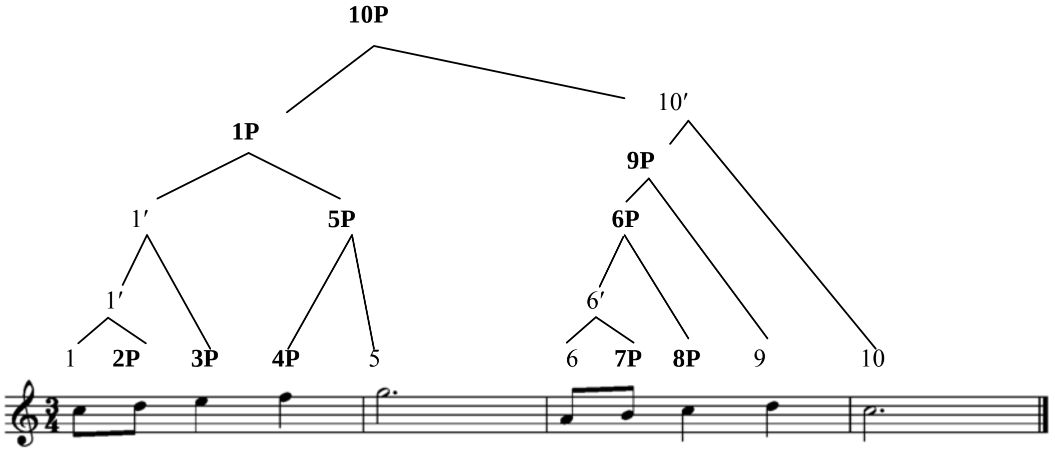
From our toy melody, a family of distinct but clearly related pieces may be derived by a simple compositional process that is crucially structure-dependent: optionally deleting the non-head of any pair of sister nodes.[14] The result in each case is felt by listeners to be a variation on the original piece that preserves its overall shape, even though certain details of this shape have been eliminated. Such variations can be called *reductions* (a notion related to the Schenkerian term mentioned in section 4.3):

---

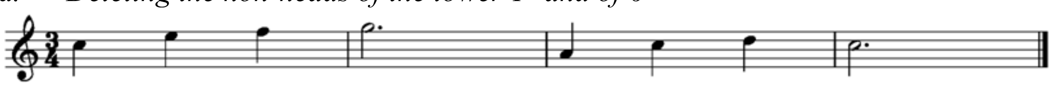
[13] Experimental results reported by Lerdahl & Krumhansl (2007) show that subjects asked to listen to a piece and judge its level of harmonic tension at various points provide judgments (even for harmonically complex music) that accord quite closely with the PR structures assigned to them by GTTM's rules, as elaborated and modified in TPS. Lerdahl & Krumhansl's subjects were musically trained as performers, but with little background in music theory.

[14] The rhythm is also adjusted to preserve its overall character, despite the loss of certain musical events. In the example given, this process is trivial, but in pieces with more complex rhythms and harmonies the details of this process raises a host of intricate puzzles. The issues are essentially the same as the "text setting problem" much discussed in recent literature (Halle & Lerdahl 1993; Hayes 1989; Hayes & MacEachern 1996; Hayes 2009; Kiparsky 2006) — for example, the fact that melodies may be adapted to fit multiple verses of a song that differ wildly in both syllable count and stress pattern (cf. Halle & Lerdahl's discussion of the shanty "Drunken Sailor"). Singers and listeners have clear intuitions that certain solutions are acceptable, while others are not.

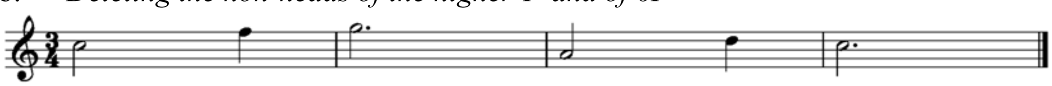
(16) **Toy melody and reductions**



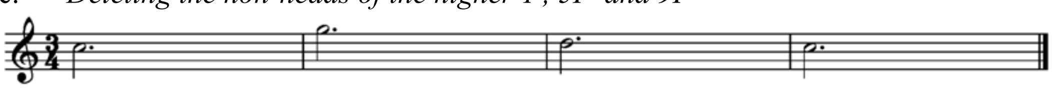
a. *Deleting the non-heads of the lower 1′ and of 6′*



b. *Deleting the non-heads of the higher 1′ and of 6P*



c. *Deleting the non-heads of the higher 1′, 5P and 9P*



Other reductions are also possible, for example one that deletes the non-heads of the lower 1′ and 6P (which would sound like the first two bars of (16a) followed by the final two bars of (16b)) or one that deletes the non-head of 1P but nothing more (which would sound like the first two bars of (16c) followed by the final two bars of the original piece). Each of these reductions is felt to be in some sense "the same piece" as the original — just simpler.[15] On the other hand, if we delete heads of constituents while maintaining their non-heads, or randomly delete elements of a melody, the result will generally not feel like a variation of the original, but rather like an entirely different piece. An example is (17). Here, just as many events have been deleted from the original as in (16c), but without regard to PR headedness. Consequently, the intuition we have about (16c) that it is in some fashion the "same piece" as the original melody, disappears in (17):

(17) **A bad reduction of (16)**

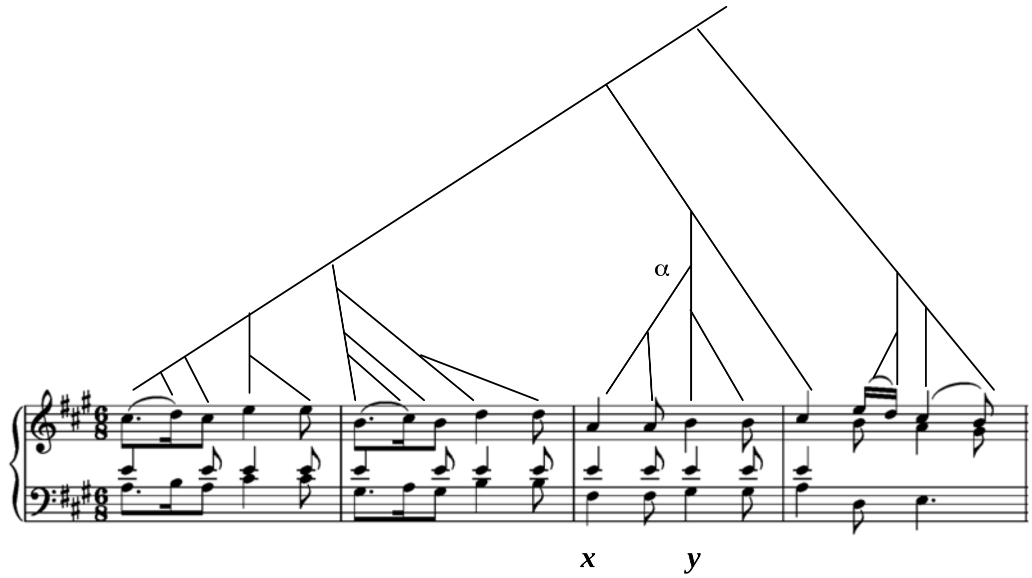
Despite the success of the PR-based procedure described above in accounting for intuitions about reductions of our toy melody, however, the matter is actually more complex. As GTTM (164 ff.)

---

[15] A variation may also be more complex than the original, so that the original theme may be a reduction of a variation, e.g.:

argues, when more complex pieces are considered, PR headedness turns out to form just one part of a more intricate picture. Consider the PR structure for the opening bars of Mozart's piano sonata K. 331 , and a low-level (i.e. not extreme) reduction that follows the procedure outlined above:

(18) **PR structure for Mozart piano sonata K. 331: GTTM notation**[16]
(GTTM 231, ex. 9.17(b)[17]; Jackendoff (1987, 231))



low-level reduction:

Now consider the two musical events labeled *x* and *y*, both of which are preserved in the low-level reduction. These two events are the heads of a pair of sister nodes immediately dominated by the node we have labeled α. The head of α is *y*, since *y* is a dominant chord and *x* is a subdominant (ignoring the suspended E, a phenomenon that we will not discuss here; but cf. Pesetsky, forthcoming). In such situations, as just discussed in connection with the daughters of 5P and 9P in (16) , it is the dominant that is chosen as head.

Suppose we now wish to produce a further reduction — one in which α, in particular, is simplified further. The PR-based procedure discussed above will dictate that if either one of the daughters of α is deleted, it will be *x*, the non-head. In fact, however, this yields the wrong result. As

---

[16] PR trees presented in GTTM and TPS indicate with hollow and filled circles those nodes whose daughters are identical (*strong prolongations*) or identical but for a different distribution of pitches among the various voices (*weak prolongations*). We omit these annotations in the PR trees presented here, including those quoted from GTTM.

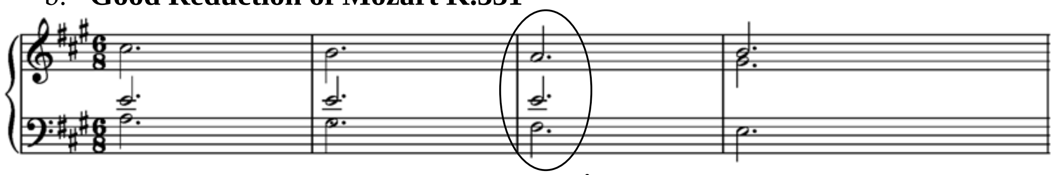
[17] GTTM's discussion of the PR representation of this excerpt entertains the possibility a second pattern of branching (taken to be a true ambiguity in the piece), which we will not discuss here, as it does not affect the points that concern us in this section.

observed in GTTM (164; which this part of our discussion summarizes), a listener strongly prefers (19b), which retains *x*, over (19a), which retains *y*, as an appropriate reduction of the original piece:

(19)a. **Bad Reduction of Mozart K.331**

preserving *y*

b. **Good Reduction of Mozart K.331**



preserving *x*

Intuitively, there are at least two reasons why *x* is retained in preference to *y* First, *x* occupies a metrically stronger position than y (first beat of the measure). In addition, retention of *x* in preference to *y* is favored by the listener's expectation that the parallelism between the first and second bars (in each of which the first chord is retained for purely PR reasons) will continue in the third.

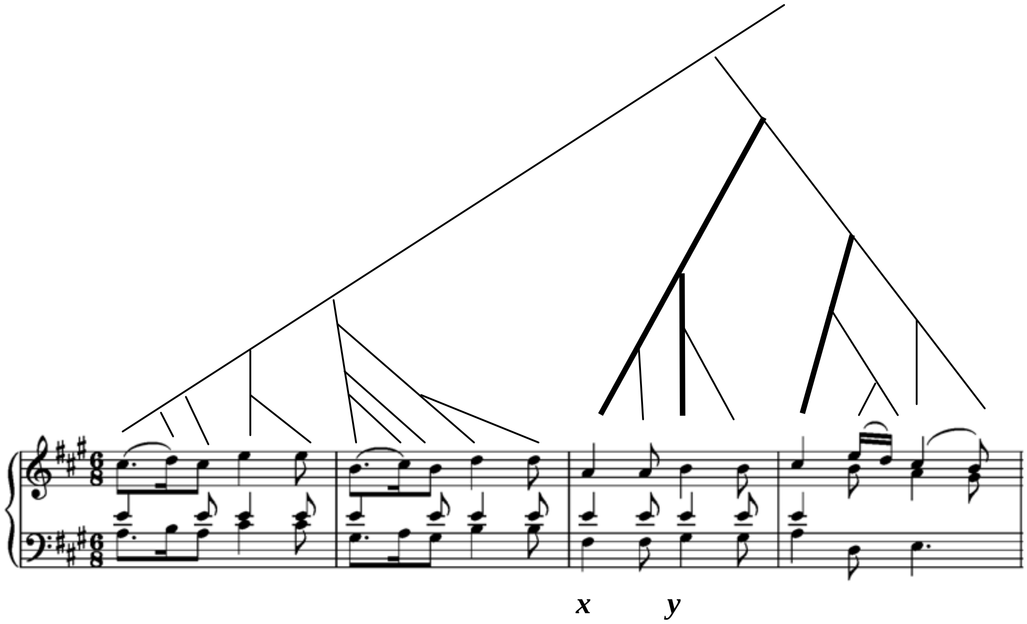
It is in response to these and similar considerations that GTTM posits the existence of TSR, a musical structure distinct from PR. Unlike PR constituency and headedness, which reflects harmonic tension and relaxation, TSR constituency and headedness reflects the hierarchical intuitions of relative *prominence* that govern our intuitions about reduction and variation.[18] Like PR, however, TSR structures are headed and binary branching — so that the reduction procedure we discussed actually applies exactly as described above, just to TSR, rather than to PR. The TSR for a given piece will differ from its PR, however, to reflect the rhythmic factors that compete with harmony to yield our intuitions about prominence.

The TSR structure for the passage in (18), as analyzed in GTTM, is (20). Note how the retention of *x* rather than *y* in (19) is reflected in the choice of *x* rather than *y* as the head of the TSR constituent that contains them both:

---

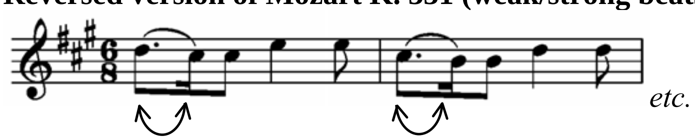
[18] Musical training does not appear to be a prerequisite for these intuitions, as an experiment by Bigand (1990; see also Bigand & Poulin-Charronat 2006) demonstrates. Two pieces that share a reduction, but do not stand in the reduction relation to each other, are also commonly taken to stand in the "variation relation" to one another (cf. Sundberg & Lindblom 1976, 111.ff). Bigand tested whether non-musicians as well as musically trained subjects could distinguish pieces related in this way — even though their rhythm and melodic contours were distinct — from pieces similar in rhythm and contour that did not stand in the variation relation to the original. Though musically trained subjects performed significantly better than non-musicians at this task, the non-musicians also performed significantly above chance. Bigand concluded that "[t]he results of the experiment confirm the general hypothesis on reduction established by Lerdahl and Jackendoff: the listener manages to hear melodies belonging to the same family as variations on an underlying pattern which is more important than the differences observed on the musical surface."

(20) **TSR structure for Mozart K. 331** **(significant deviations from PR boldfaced)**



Crucially, although such rhythmic factors as metrical strength, spacing among events, and parallelism all play a role in building TSR, the harmonic factors relevant to PR also play a role in constraining TSR. Though metrical strength overrules harmonic tension in picking *x* as more prominent than *y* for TSR purposes, in other cases, harmonic factors outweigh rhythmic factors. For example, if the first and second chords of bars 1 and 2 of Mozart K. 331 are reversed, our intuition about reduction remains unchanged — despite the fact that it is a weak beat that is being retained. A version of the sonata that begins as in (21) would in fact have the same reductions as the version that Mozart actually wrote:

(21) **Reversed version of Mozart K. 331 (weak/strong beats exchanged)**



Here, the expectation that pieces begin with a tonic chord appears to overrule metrical strength in determining perceived prominence in the first bar, and parallelism with the first bar plays the same role in the second bar.

The default thus seems to be isomorphy between PR and TSR, with rhythmic factors permitted to create non-isomorphy under particular circumstances. GTTM's theory of the PR-TSR mapping thus contains two components. First, GTTM offers a formal theory of permissible and impermissible non-isomorphy between PR and TSR, independent of the factors that might motivate non-isomorphy. In addition, GTTM proposes a set of violable constraints (*preference rules*) that motivate or inhibit these discrepancies. We will not be discussing the violable constraints here, but focus instead on the first part of their proposal — the formal theory of permissible discrepancies between PR and TSR.[19]

---

[19] Lerdahl and Jackendoff's preference rules, an innovation that plays a key role throughout the GTTM system, famously anticipated Optimality Theory (Prince & Smolensky 2004, 1993). If optimality-theoretic interactions among constraints are characteristic of language as well, as has been widely (if

4.6 Mapping from TSR to PR (music)

In our presentation so far, PR has been the focus of discussion, and we have introduced TSR as if it were an emendation derived from PR. As described in GTTM (pp. 216ff.), however, it is actually PR that is formed from TSR — by a top-down procedure.[20] We present here a "realigned" (and slightly simplified) version of that procedure, which relies on a notion that we will call ***RD* ("root distance") *number***[21]

(22) **RD number**
The *RD number* of an event *e* in a structure K, RD(*e*), is the number of nodes that non-reflexively dominate the maximal projection of *e* (i.e. *e*P) in K.

This is where the notion "maximal projection" turns out to be crucial to the syntax of music, as promised in the final paragraph of section 4.3, much as it is to language. The mapping procedure itself proceeds stepwise by RD level, forming optimally stable harmonic connections among the events at any given level — with an 'exception clause' that permits reference to one additional level when the result enhances harmonic stability. An explicit set of stability conditions are given in GTTM (p. 224). After connections are made at one RD level, the algorithm proceeds to the next lower level. The notion "stability of prolongational connection", for our purposes, can be equated with perceptual proximity in a tonal pitch space between two chords attached as PR sisters.[22]

To see how GTTM's procedure works in practice, consider once more the relation between the TSR and PR structures for the opening of Mozart's K.331 sonata. We repeat GTTM's TSR, with relevant RD numbers added below:

---

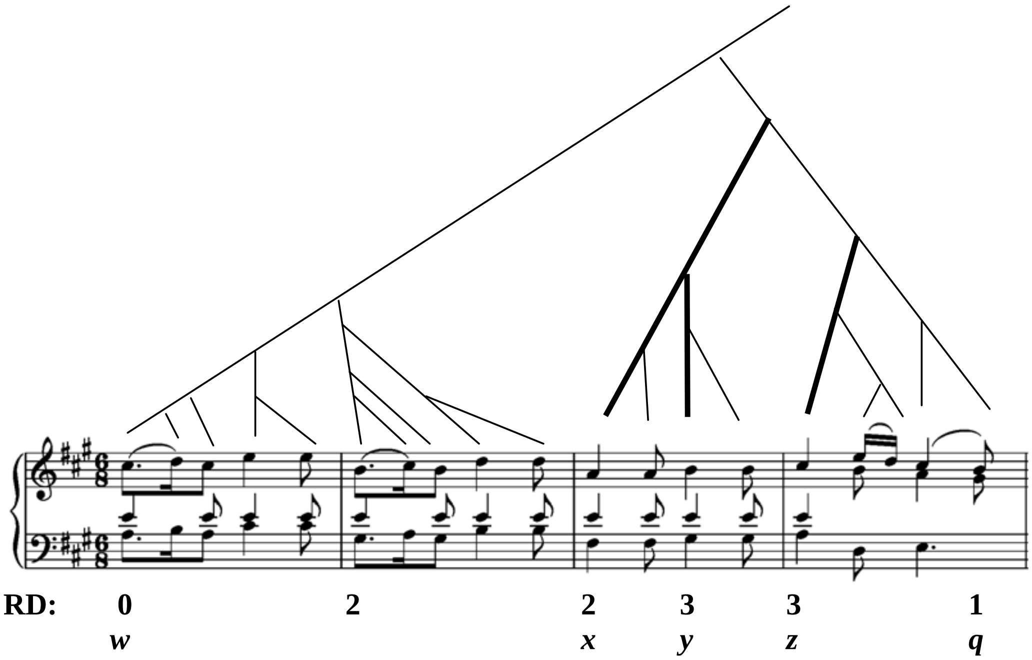
controversially) argued over the past two decades, this too might be a fruitful area in which to explore consequences of the Identity Thesis in future work. Katz (2007), for example, argued that the content of GTTM's Well-Formedness Rules on Grouping actually falls out of GTTM's own Preference Rules, if these are recast in an explicitly optimality-theoretic setting.

[20] The "official boxology" of GTTM, as discussed above, posits no directionality to the mapping, but their detailed discussion of PR and TSR consistently derives the former from the latter.

[21] The term "RD number" is ours, but the concept is implicitly invoked throughout GTTM's description of the PR-TSR mapping. Given this top-down procedure, the TSR prominence of each musical event is decided at the point where the procedure first encounters a phrase headed by that event — i.e. at that event's maximal projection.

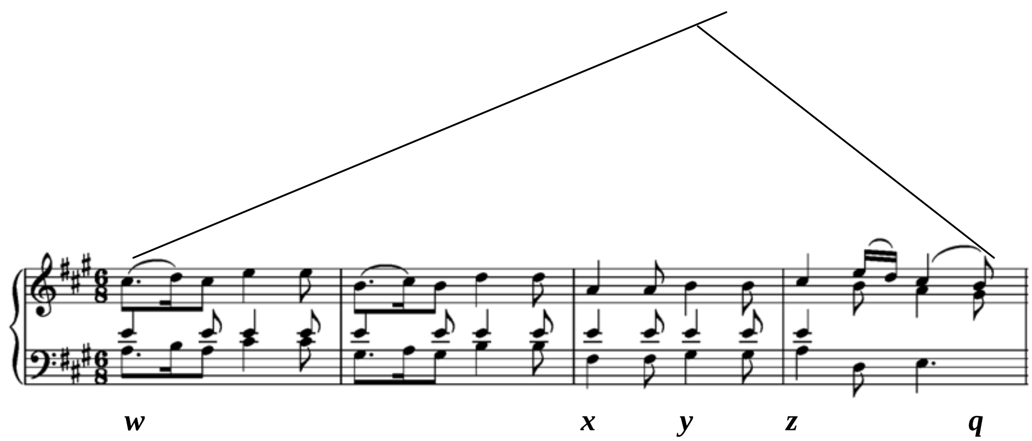
[22] For a full justification and formalization of this notion, see Krumhansl (1990) and TPS. Both GTTM and TPS note a variety of other factors that play a role in prolongational stability, which we ignore here.

(23) **TSR structure for Mozart K. 331** **(significant deviations from PR boldfaced)**



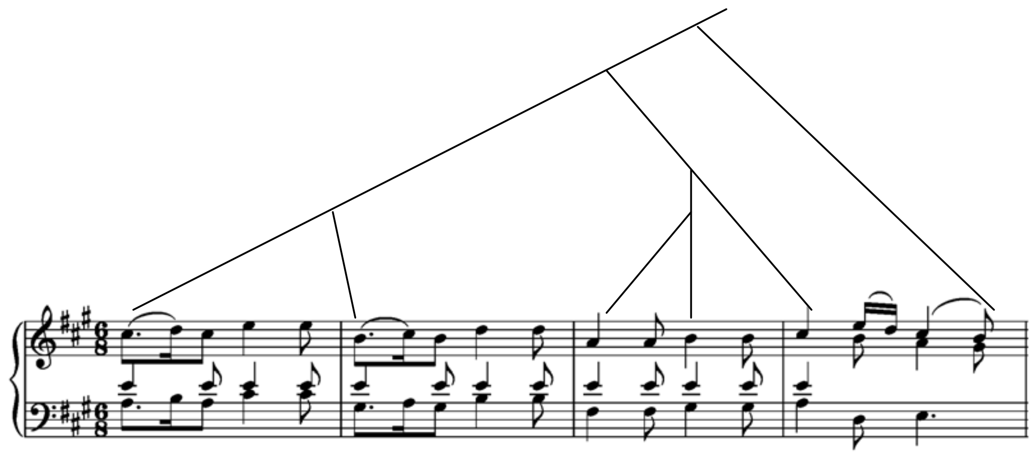
The first step of the procedure will attach *q* as a right-sister to *w*, since RD(*q*)=1 and RD(*w*)=0:

(24) **Step 1 of the GTTM procedure**



The next step considers those events whose RD number is 2 and asks how they should be included in the structure (24) that was just built. If there were no exception clause, *x* would either attach as a left-sister to *q* or as a right-sister to *w*, and the procedure would advance to the next step. Instead, however, the exception clause mentioned above is invoked, since a much more stable set of configurations can be created by "reaching down" one RD-level and attaching some events whose RD number is 3 (*y* and *z*) — as part of the same step that considers possible attachment sites for *x*. In particular, if the constituent [*z w*] is created, it will consist of two identical chords (a *strong prolongation*, in GTTM's terminology) — the most stable possible configuration. If the constituent [*y z*] is created, it will consist of a chord built on the fifth degree of the scale (the *dominant*, see section 5) followed by a chord built on the first degree of the scale (the *tonic*) — also an extremely stable configuration. Since both configurations are more stable than any alternative in which *x* is attached as a sister to *w* or *q*, the exception clause creates these configurations.[23] Chord *x* is then attached as a right-sister to *y* (forming a subdominant-dominant chord progression, also relatively stable). The result is the partial PR-structure given in (25), from which the full structure shown in (18) can be derived by reapplying the procedure as required:

(25) **Step 2 of the GTTM procedure**



RD: 0 2 2 3 3 1

w x y z q

In more complex examples, other factors discussed in GTTM help to resolve attachment ambiguities. At each stage in the process, if two or more candidate PR structures are equally stable, other considerations (which take the form of Preference Rules; see n. 19): resolve the tie.

The top-down procedure thus results in a PR structure that may differ in constituency and embeddedness from the corresponding TSR structure. The restriction of each step to a single RD level plus a limited exception clause, however, ensures that corresponding elements in the two structures will, more often than not, be equally embedded (cf. GTTM's "Prolongational Hypothesis 2", p. 213). TSR-PR isomorphy may thus be viewed as a kind of default.[24]

4.7 Mapping from syntax to prosody (language)

The mapping from syntactic to phonological domains in language has been the subject of much research in generative linguistics, beginning with the work of Selkirk (1972, 1984). Here we consider a simplified version of Selkirk's (1996, 2000) proposals. In this model, though prosodic structure is derived from syntactic structure, it is independent and not isomorphic with it.

The most important feature of the system for the present discussion is that edges of certain prosodic categories are aligned with the edges of certain syntactic categories. For example, in English,

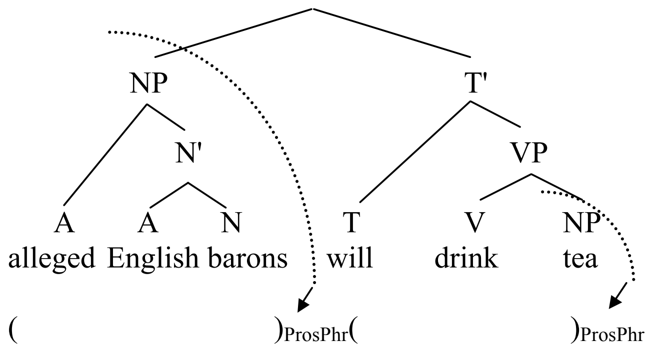
---

23 In TPS (pp. 16-17, 159-161), Lerdahl proposes a severe restriction on the ability of the TSR-to-PR mapping procedure to look at the next RD-level, limiting it to configurations in which the result is a right-branching strong prolongation. We will continue to assume the looser proposal from GTTM here, since the tighter proposal excludes numerous TSR-PR disparities that seem to us to exist — such as the one from GTTM's analysis Mozart's K. 331 discussed here.

24 There is also a global influence from a schema that GTTM calls *normative prolongational form*, which we touch on in section 6.

the right edge of each NP that bears stress[25] is aligned with a prosodic phrase boundary. In language, as in music, one can get a rough idea of whether a group boundary is appropriate by placing a prolonged pause there and assessing the felicity of the pause. As we will see, there is a wide variety of other boundary diagnostics that have been proposed cross-linguistically.

(26)



Prosodic headedness (prominence) is assigned based on phonological or lexical properties of the constituents. In English, word-level stress is partly predictable by rule, and partly idiosyncratic — so certain aspects of word-level stress must be stored in the lexical entry of a word (i.e. memorized). At levels above the word, head prominence in English is generally assigned to the rightmost stressed syllable in each constituent (although important exceptions exist). Some phonological processes are conditioned by the presence or absence of prosodic boundaries. Most researchers agree that the phrase boundaries of syntax play no direct role (but cf. Kaisse 1985 for a different view).

4.8 Aligning the two systems

On the surface, there is not much similarity between the place of syntax in the linguistic model and the place of PR in the GTTM model. Language does not seem to have anything like the mapping from TSR to PR, and music does not look as though it has a syntax that maps to phonology. This might look like a challenge to the Identity Thesis. Music and language might just differ in this domain, and the Identity Thesis might be too strong. Alternatively, however, the differences might be more apparent than real — artifacts of how the proposals are presented rather than properties of the proposals themselves. In this section, we pursue the second possibility.

In particular, we argue that many of the apparent differences between music and language in this domain arise from the way in which proposals have been presented, rather than their actual content. GTTM describes the interaction between TSR and PR in terms of a procedure that maps TSR into PR. If we align TSR with linguistic prosody and PR with linguistic syntax, then GTTM's presentation of the mapping is, of course, the opposite of what linguists generally assume (i.e. from syntax to prosody).

Within language, there is indeed a sense in which syntactic structure is privileged over prosodic structure: prosodic structure can (mostly) be recovered from syntactic structure, but not vice-versa. The privileged status of syntactic structure has often been described in terms of a derivation, in which some, but not all, information from the syntactic-structure input determines the prosodic-structure output (Nespor & Vogel 1986, Selkirk 1984, Hayes 1989). More recently, however, it has been argued

---

25 NPs that do not normally bear stress (e.g. weak or clitic pronouns) are not generally mapped to prosodic edges.

that the very same observations may be captured by non-derivational theories, with systems of constraints that filter a larger (possibly infinite) set of candidate syntactic-structure/prosodic-structure pairs (Selkirk 2000, Truckenbrodt 1999).[26] The privileged status of syntactic structure is, on this view, a property of the constraint-set.

On the music side, we suggest that it might be profitable to ask questions similar to those that have been central to linguistic discussion. Not only does GTTM propose a general directionality of mapping between TSR and PR that has no parallel in linguistic research, it also posits a specific "top-down" orientation to this mapping that has not been argued for language. Can the insights of this model be equally well described non-derivationally? If so, differences in directionality will no longer provide any intrinsic obstacle to the alignment of TSR with prosody and PR with syntax. What will be important is the nature of the constraint set and the representations to which they apply. Are *they* the same in language and music? In what follows, we argue that a non-derivational characterization of the relationship between TSR and PR is possible, and cautiously suggest that the constraints necessary for music and language are indeed the same.

We first recast the TSR-PR mapping in terms of global constraints on the correspondence between the two components, removing directionality from the theory. This allows us to ask whether the "boxology" and direction-of-scan differences between GTTM's approach to the PR-TSR mapping and the various approaches to the syntax-prosody relationship reflect genuine differences between music and language. We find that one might understand both mappings as equally top-down, bottom-up, or direction-neutral — without complicating the conception of linguistic or musical structure.

We make the claim here that both the syntax-prosody mapping in language and the TSR-PR mapping in music are governed by an overarching constraint stated in its general form in (27):

(27) **Region Conditions for music and language: General form**
For every pair of distinct events (α, β), such that α and β have property *P*, if *e* linearly intervenes between α and β, both α and β exceed *e* in prominence.

What differs between language and music is the nature of the property *P* referred to in the region condition. In this section, we show how this general constraint can be used to characterize both formal systems.

4.8.1 The TSR-PR correspondence

The first step in our endeavor is to note that the derivational property of the procedure with which GTTM derives PR from TSR — and *a fortiori*, the top-down character of the derivation — does no actual work within the theory. What matters is only the notion "distance from the root" (i.e. RD-number). To see this, consider (28), where the subscripts indicate RD-number (i.e. TSR level). Let us assume that each event in (28) is unique, to keep prolongation out of the picture.

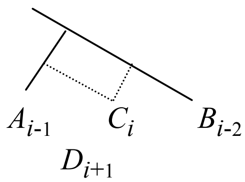
In (28), *A* has been attached as a left-elaboration of *B* at the previous derivational level i-1. At the current level *i*, the system decides whether *C* becomes a sister to *A* or to *B*. Finally, *D* is "waiting in the wings" to be attached at the next derivational level i+1. As far as we can tell, the decision about *C* that the system makes at the current level *i* has no consequences for the decision concerning *D* that it

---

[26] This approach was developed within Optimality Theory, but a general approach along these lines could be developed in other frameworks as well.

will make at the next level i+1 . Nothing whatsoever seems to hinge on the fact that the GTTM/TPS system attaches the sister of *C* before it attaches the sister of *D*:

(28)



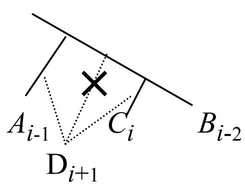
What *is* important to the system is a different fact about TSR-level. Putting aside the "exception clause" for a moment, we observe that *D* may attach as a PR sister of any element whatsoever in (28), so long as the following (preliminary) Region Condition for Music is obeyed:

(29) **Region Condition for music (cf. GTTM, 217-219; also TPS 15)**
For every pair of distinct PR events (α, β),
such that α and β are the heads of sisters ["Property P"],

if an event *e* linearly intervenes between α and β, both α and β exceed *e* in prosodic prominence (i.e. RD(α)<RD(*e*) and RD(β)<RD(*e*)

Given (29), the PR sister assigned to *D* in (28) may be *A* or *C*, but it may not be *B*, since *C* is more prominent than *D* and linearly intervenes between *D* and *B*:

(30)



Thus, to know whether the attachment of an event *α* to any other event *β* is well-formed, it is necessary to know only *whether any element that is closer to the root than α or β linearly intervenes between them.* In (30), for example, *C* is closer to the root than *D* is (and linearly intervenes between *D* and B). That is why the attachment of *D* to *B* violates the Region Condition. The exception clause that was crucial to the PR-TSR disparity in Mozart K.331 can then be incorporated straightforwardly:

(31) Exception to (29)
...unless RD(β)-RD(*e*) ≤ 1 and β forms a more stable prolongational connection with *α* than *e* does

At this point, it should be clear that the sequential property of the top-to-bottom procedure is not necessary. It is true that in the derivation posited in GTTM, events closer to the root than some element α are easily distinguished by the fact that they have already been integrated into PR structure while α itself has not — and that events *further* from the root than α (putting aside the exception clause) are not yet visible at all. Because it is important to distinguish events closer to the root from events further from the root, and because the derivational procedure makes this distinction, it might have seemed that the derivational, top-down property of the GTTM theory was crucial. In fact, however, the top-down derivational procedure plays no active role in the theory beyond encoding RD. This fact will be useful when we compare TSR to linguistic prosody.

4.8.2 The syntax-prosody correspondence

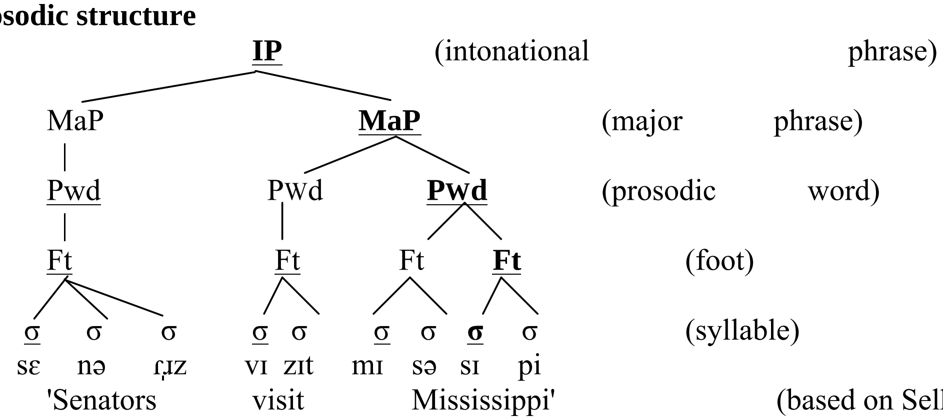
GTTM's references to TSR levels, i.e. RD number, embody a hypothesis: that an event's *prominence* correlates with the distance from the root of its maximal projection in TSR. The precise denotation of "prominence" for music involves a complex interplay of rhythmic and harmonic factors, as discussed in GTTM (pp. 117-119). There are, of course, phenomena in the prosody of *language* that are also described in terms of *prominence*, a term whose denotation is a similarly complex research topic. What is its structural correlate?

As we saw above, the notion of property *P* from the generalized region condition in (27) relevant to music is syntactic (PR) sisterhood. If the identity thesis with which we started this paper is correct, we predict that any difference between language and music in this regard is due to the presence of a lexicon in language and its absence in music. In this section, we will indeed explore such a difference — which will indeed turn to hinge on the lexicon.

In Selkirk's framework, the phonological "events" of a sentence are recursively parsed into hierarchically organized groups, e.g. ***utterance, intonational phrase, major phrase, minor phrase*** (possibly not present in English)**,** and ***prosodic word***. Within each group, one subconstituent is the most prominent, and constitutes the *head* of the group (the property called *culminativity*).[27]

So far, linguistic prosodic structure looks just like TSR in music, except that non-terminal nodes in linguistic prosodic structure are identified by particular *prosodic designations* (our term) drawn from the list of group types discussed above:

(32) **Prosodic structure**



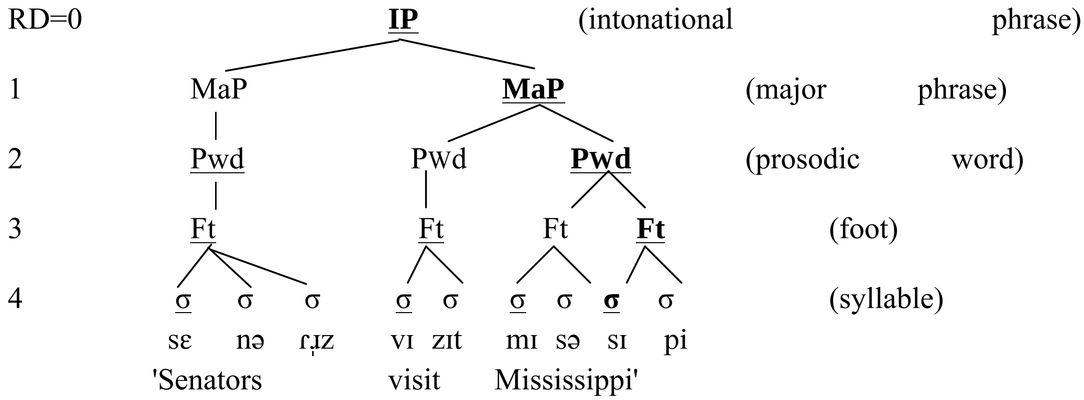
(based on Selkirk (2008))

If this proposal is correct, language appears to make fewer prominence distinctions than music does (according to GTTM). For the moment, let us accept this as a real difference, but we return to this topic below.

RD number correlates with prominence in both systems. To see how RD number could be understood as expressing prominence in linguistic prosody, consider (33):

[27] We will not discuss here the possibility of cross-linguistic variation in *culminativity* (Hayes 1995, p. 24), which could, of course have implications for variation among musical systems.

(33) **Prosodic structure (showing RD numbers)**



In (33), the most prominent element is the syllable [sɪ] in *missi-SSI-ppi*, with RD number 0; the next most prominent element is the syllable [sɛ] in 'SE-nators', with RD number 1; and so on. Each RD number corresponds to a particular prosodic designation. For instance, *major phrase* corresponds to RD number 1 and *syllable* (notated as "σ") corresponds to RD number 4. The designations, in essence, are just names for the RD numbers. Note that the hypothesis for language according to which there is a limited number of prosodic designations amounts to the hypothesis that depth of embedding is strictly limited.

The reason linguistic work in this area has converged on representations labeled with prosodic designations as in (33) is the observation that phonological and phonetic processes often target particular prosodic units, and not others. For example, Russian and Polish have a process that changes a voiced obstruent consonant to its voiceless counterpart (e.g. *b* to *p*). This process affects only the final consonant in a Prosodic Word. The Bantu language Chi-mwi:ni, in turn, has a process that shortens a long vowel — but only at the end of a Major Phrase (Kisseberth & Abasheikh 1974, as analyzed in Hayes 1989).

Phonologists thus posit specific distinct prosodic designations as a response to the fact that certain processes are sensitive to the domains corresponding to those designations. The presence of boundaries between various prosodic constituents is determined with reference to the syntactic features of the lexical items appearing in an utterance, which we discuss in the next section. In other words, the distinction between various prosodic categories depends in part upon lexical information. The fact that these distinctions appear not to be present in music, then, is predicted by the identity thesis.

To summarize the discussion so far: some of the most obvious distinctions between how TSR is presented in GTTM and how prosody is discussed in the linguistic literature have turned out to be illusory (as in our discussion of RD number and top-down derivations). Other distinctions appear to be the result of the presence of a lexicon in language but not music, and are thus consistent with the identity thesis. The discussion has not provided an argument specifically *in favor* of the alignment of PR with syntactic structure and TSR with linguistic prosody. Instead, we have shown that a number of putative obstacles to this alignment are not in fact obstacles at all. In the next section, we discuss some stronger evidence for the alignment: specific details of the mapping between syntax and prosody that are the same in music and language.

4.8.3 Region conditions

We have seen in the previous sections that there is a law-governed grammatical relationship between syntactic structure and prosodic prominence in both language and music, and that the formal nature of syntactic and prosodic structures is identical in these domains. In this section, we argue that the mapping between syntax and prosody might also be identical. Much as the Region Condition in music requires a decrease in TSR prominence between the heads of PR-sisters, so the linguistic relation between syntax and prosodic structure requires a decrease in prosodic prominence between certain designated points in syntactic structure ("Property P"). If both music and language invoke Region conditions with the general form (27), this fact alone supports the specific parallels between music and language that we have been arguing for.

The devil in the details is the nature of "property *P*". In music, as we have discussed, α and β are the heads of sister constituents. For language, however, we have not so far specified the nature of "Property *P*". if Selkirk's work and related research is correct, α and β are determined by the properties of specific lexical items, and not syntactic sisterhood. At first glance, this appears to constitute a substantive difference between language and music. At the end of this section, however, we will suggest that the characterization of "property *P*" relevant to language might actually do no harm if inserted into the Region Condition relevant for music as well (and that the characterization relevant to music might also do no harm when applied to language). The key factor is the absence of a lexicon in music. If this is correct, we have an even stronger argument for the Identity thesis, since both music and language involve a single Region Condition.

In language, there are two "properties *P*" that appear to be relevant to the Region Condition. The first is the property of being the *linearly rightmost element* of a kind of syntactic maximal projection called a *phase*. We will call elements that occupy such positions ***phase markers.*** Noun Phrases, full clauses and certain types of Verb Phrases are believed to function as phases — but we will limit our discussion of phases and phase markers to Noun Phrases. Phases have an importance in current syntactic theory that extends beyond prosody. They are the syntactic units whose construction by Merge triggers the transfer of syntactic information to both semantics and phonology, and they play a wider role in limiting the possibilities for syntactic movement (Internal Merge) and morphological agreement. A second "property *P*" relevant to the Region Condition for language is the status of a terminal element as a ***clitic or non-clitic***. A clitic is a lexically unstressable word that is dependent on a stressable host. With these notions in place, we can state the Region Condition for music and for language as in (34), with the notion *prosodic prominence* defined as in (35):

(34) **Region conditions for music and language**
For every pair of distinct syntax/PR events (α, β), such that

a. **Relevant to music:** α and β are the heads of sisters; or

b. **Relevant to language:** α and β are phase-markers, and β is the phase-marker linearly closest to α; or

c. **Relevant to language:** α and β are non-clitics, and β is the non-clitic linearly closest to α,

if an event *e* linearly intervenes between α and β, both α and β exceed *e* in prosodic prominence.[28]

Exception (*music*)*:* ...unless RD(β)-RD(*e*) ≤ 1 and β forms a more stable prolongational connection with *α* than *e* does

(35) **Prosodic Prominence**
α exceeds β in prosodic prominence iff RD(α) < RD(β).

The Region Condition fulfills similar goals in both language and music: constraining non-isomorphy between prosodic (TSR) and syntactic (PR) structure, and encoding some information about phrase boundaries in the form of prominence distinctions. The existence of such a condition in both domains supports the Identity Thesis in a general fashion. The differences seen in (34a) vs. (34b-c), however, also pose a clear challenge for the thesis. In music, the nature of α and β is determined by syntactic sisterhood, while in language it is non-clitic status and location at the right edge of a phase that matter. It might seem that we must be dealing with two formally distinct cognitive systems.

In fact, however, the Identity Thesis already acknowledges one important difference between language and music: the *basic ingredients* that are combined recursively by Merge: lexical items in language vs. pitch classes and pitch-class combinations in music. The challenge for the Identity Thesis is to show that all apparent differences between the two domains can be derived from this fundamental distinction.

Consider (34c) in this light. The property of being a clitic is a *lexical* property. Cross-linguistically, it is closed-class function words that tend to be lexically marked as clitics, but the actual list of words that have this property differs from language to language, and even with a given language, near synonyms may differ in clitic status. If music lacks a lexicon, as we have suggested, then the conditions under which (34c) applies will never be met in music. It is thus completely possible that the overall grammar of music formally includes (34c) (because the grammar of music and the language are the same), but never uses it — because music has no lexicon and thus has no clitics or non-clitics.

Very similar considerations are relevant to (34b). The status of a maximal projection as a phase (or non-phase) depends on its head's part of speech. If current proposals about phases are correct (Chomsky 2001, 2004), for example, TP is not a phase, but CP is a phase. Part-of-speech information is an attribute of lexical items. Consequently, it is once again possible to imagine that the grammar of music formally includes (35c), but never happens to use this condition in constraining the PR-TSR

---

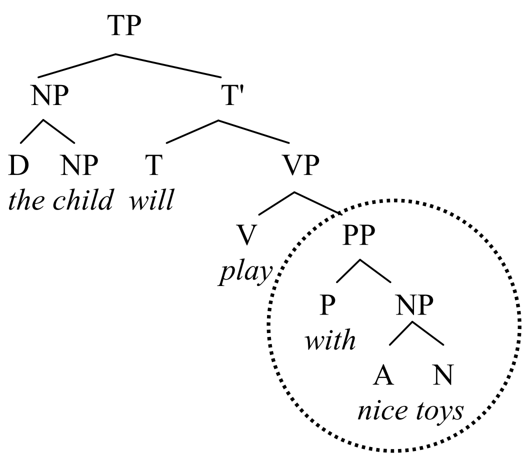
[28] The alternative, "*e* does not exceed either α or β in prominence" would not prohibit in language a prosodic structure in which heads of PW and MajP are equally prominent — presumably a bad result.

relation — for lack of a lexicon. Needless to say, similar reasoning applies to the exception clause in (34), which makes crucial reference to the concept of "stable prolongational connection". This concept has to do with harmonic distance and other relations proper only to music, and irrelevant to language.[29]

Note that the point of this discussion is not to saddle music *a priori* with a large set of unusable conditions that make reference to linguistic notions like part of speech and clitic, nor to saddle language *a priori* with a set of equally unusable conditions that make reference to harmonic distance and pitch class. Rather, we are making an *observation*, which, if correct, supports the Identity Thesis: those conditions that seem different in language and music all make crucial reference to the building blocks of the two domains — which, as we have repeatedly stressed, *are* different. It is those properties of language and music that are purely *formal* that do not appear to differ.

Note, however, that one apparent obstacle to this view still remains: clause (34a) of the Region Condition, which identifies α and β by PR sisterhood. Can this condition — which is crucial to music and does not make crucial reference to music-particular building blocks — be said to hold for language as well? On a first pass, (34a) seems problematic for language, since it may make erroneous predictions about left-branching syntactic structures. Consider (36):

(36)



The circled complement PP is problematic for this approach. The preposition *with* and the noun *toys* are heads of constituents that are syntactic sisters. Region condition (34a), therefore, requires that any element intervening between them (*nice*, in this case) be less prominent than both. This is not the pattern of prominence that we observe in productions of sentences like (36), however. In fact, the preposition *with* is generally unstressed, while *nice* is generally stressed, which violates (34a). Consequently, it might seem that any extension of (34a) to language makes the wrong predictions — a conclusion that would be unexpected under the Identity Thesis.

Notice, however, that the very pronunciation of (36) that is favored by (34a) actually violates (34c). The verb *play* and the adjective *nice* are non-clitics — and *nice* is the non-clitic linearly closest to *play* . Condition (34c) requires that any element intervening between them (*with*, in this case) be less prominent than both. Conditions (34a) and (34c), then, are in conflict here. Condition (34a) requires that *nice* be less prominent than *with*, but (34c) requires that *with* be less prominent than *nice*. If the grammar resolves the conflict so that condition (34c) takes precedence over condition (34a), the theory

---

[29] Intriguingly, however, Lerdahl (2001a; 2004) has suggested that poetic rhyme is a linguistic form of GTTM's notion of *strong prolongation*, with strong formal similarities to its musical counterpart. See also Katz (2008) for an extension of these ideas to hip-hop.

now makes the correct prediction about prominence in examples like (36)   Because most sentences contain a rich mix of clitic and non-clitic lexical items, situations like (36) in which satisfaction of (34a) gives way to satisfaction of (34c) will be common.[30] On the other hand, in music, where (34c) never applies (for lack of lexical items), the effects of (34a) will be pervasive.

If the conclusions of this section are correct, there remains no reason to claim that any of the Region Conditions for music and language are *formally* different. The absence of a lexicon in music and its presence in language can account for the apparent predominance of (34a) effects in music and its lesser role in language, as well as the total absence of (34b-c) effects in music. We may thus conclude that the Region Conditions governing the mapping between syntactic structure and prosodic prominence are identical in music and language. If this is so, then the interface conditions linking syntactic and prosodic structure are the same across the two domains, with apparent differences all derivable from the presence of a lexicon in language and its absence in music — exactly as posited by the Identity Thesis.

5 Internal Merge and the Cadence

5.1 Internal Merge in Language

The presentation of Merge in section 3.3 presupposed some fairly straightforward assumptions about the relationship between hierarchical structure and the semantic combinatorics of lexical items — and equally straightforward assumptions about the relationship between hierarchical structure and the phonological surface string. On the semantic side, for example, the fact that the verb *read* in *The girl will read the book* merges with *the book*, rather than with some other element (e.g. *will*) reflects the fact that the meaning of this noun phrase semantically combines with the meaning of the verb, supplying the predicate with one of its needed arguments. On the phonological side, we assumed that the linearization of the terminal elements of a structure produced by iterated recursive Merge should be equally straightforward: when two elements merge, the right edge of one element aligns with the left edge of the other.

---

[30] It might be possible to see the effects of (34a) in language as well, whenever a configuration like the VP of (36) lacks clitics. Suppose we adopt the widely accepted (but not uncontroversial) proposal that it is the determiner (e.g. *the* or *each*) — not the noun — that heads those phrases we have been labeling as NP. On this hypothesis, the VP of sentences like (i)-(ii) will have the structures indicated:

(i) The child will [$_{VP}$ visit [$_{DP}$ the [$_{NP}$ big city]]
(ii) The child will [$_{VP}$ visit [$_{DP}$ each [$_{NP}$ big city]]

On the assumption that *the* is a clitic, but *each* is not (though both belong to the category D), example (ii) illustrates the point at hand. Since *each* and *city* are the heads of sister nodes (D and NP, respectively), (34a) predicts that the intervening adjective *big* should bear less prosodic prominence than *each* and *city* — and (34c) is irrelevant. This prediction seems to us to be correct. When example (ii) is uttered in a neutral context, the most natural pronunciation shows a dip in prominence between *each* and *city*. By contrast, in (i), where the determiner *the* is a clitic, (34c) requires *the* to be less prominent than *big* (and less prominent than *visit*) — though (34a) will assign greater prominence to *city* than to *big*. Once again, this prediction appears to be correct. Thus, there might be some reason to conclude that (34a) does apply in language, even though its effects are often masked.

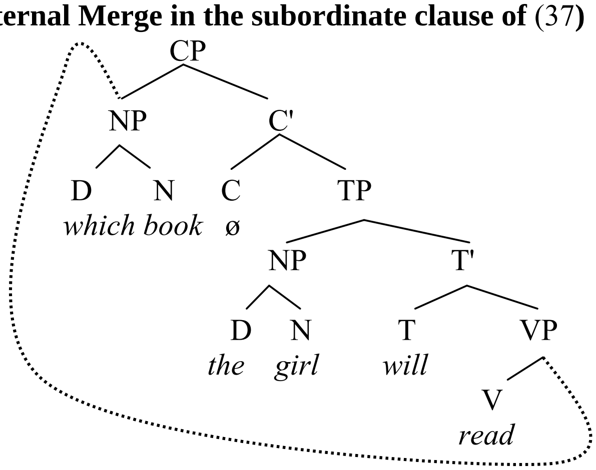
Because of the straightforwardness of these assumptions, considerable attention has been devoted to cases in which the otherwise predictable relation between hierarchical syntactic structure and the phonological surface breaks down, such as the bracketed subordinate clause in (37):

(37) The boy wonders [which book the girl will read __].

In this example, the nominal ***wh*-phrase** *which book* is the direct object of the verb *read*, and thus should have merged with the verb *read* — yielding a pronunciation in which *which book* follows (or is at least adjacent to) *read* (as indicated by the underscore). In fact, however, *which book* is pronounced at the left edge of its clause. According to our assumptions about linearization, this means that *which book* must have merged with some element to the left of (and structurally higher than) the verb. Examples like (37), of course, instantiate a well-known class of constructions in which an otherwise expected relationship breaks down between a semantically relevant principle of syntactic structure ("a verb merges with its direct object argument") and a phonologically relevant principle ("elements that merge are pronounced contiguously").

For such constructions, substantial evidence favors a particular resolution of this problem: the *wh*-phrase in examples like (37) has actually merged in more than one position, with the phonology choosing to pronounce the phrase in only the highest of these positions (a point on which languages differ, of course; cf. Huang 1981). This possibility of merging an already-merged element a second time (*syntactic movement*), has been dubbed *Internal Merge* by Chomsky (1995b). When we first discussed and demonstrated Merge in (5)-(6), the elements that participated in Merge were always independent of each other: either lexical items or independent phrases (sets) produced by previous applications of Merge. Merge of this sort may now be called *External Merge*. As Chomsky suggests, *Internal Merge* is just the third logical possibility (perhaps even an *expected* possibility; cf. Chomsky 2005, 12-13), in which one of the two Merging elements is a subpart of the other:

(38) **Internal Merge in the subordinate clause of** (37)



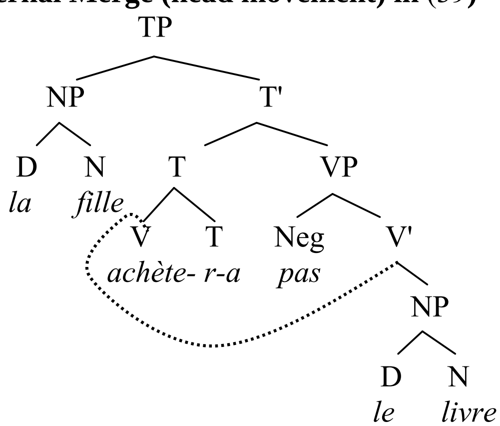
Wh-questions like (37) show Internal Merge applying to a complex phrase (*phrasal movement*), i.e. Internal Merge of an element that was itself produced by Merge. Somewhat more controversially, it has also been argued that Internal Merge may also move the lexical heads of larger phrases without displacing the phrase itself (*head movement*). It is in fact head movement that will concern us most for the remainder of this section, and provide the basis for our continued discussion of the Identity Thesis. Consider in this light the French sentence in (39):

(39) **French finite clauses: V - Tense - Negation - direct object**
la fille (n') achète+$[_{T}$ r -a] pas le livre.
the girl *prt* buy -will-3SG not the book
'The girl will not buy the book.'

This example presents a problem for straightforward views of phonological linearization reminiscent of the problems posed by (37). The finite verb 'buy' takes a direct object, which we expect to undergo Merge directly with the verb itself, producing a VP headed by that verb. We expect the VP to be pronounced with the verb and its object contiguous, just as in English. Instead, however, the verb is pronounced to the left of Tense — and seemingly merged with it, especially since the verb and Tense are *tightly coupled*, forming a phonological word. The verb itself is separated from the rest of its VP by the expression of Tense, and is certainly not adjacent to its direct object. Not only Tense, but the negation also intervenes between V and its object. Just as in (37), the link between syntactic structure and linear order has broken down.

Internal Merge once explains the discrepancy. In an example like (39), the verb *acheter* 'buy' did Merge with its complement *le livre*, as expected by the normal laws that link thematic role assignment to syntactic position. It then *merged a second time* with T:

(40) **Internal Merge (head movement) in** (39**)**



This example highlights several properties of head movement that will figure prominently in the following discussion: [31]

---

[31] Head-movement phenomena differ from phrasal movement with respect to properties (41a) and (41d), among others. In response, Matushansky (2006) has suggested that head-movement *per se* does not have these properties, but that head movement is followed by a separate process of "*m*[orphological]-merger" that tightly attaches the moved head to its host. This view is probably compatible with the proposals concerning music made in this paper. Alternative proposals in the syntactic literature that attempt to entirely dispense with head movement as a species of Internal Merge, such as the suggestions of Mahajan (2000) and Koopman & Szabolcsi (2000) that achieve the linear effect of head movement with a chain of phrasal movements are probably not compatible with our proposals for music, but we will not develop this point here.

(41) **Characteristic properties of head movement in linguistic syntax**

a. Once the head H of a phrase HP has undergone head movement, H is **pronounced string-adjacent to the head of a higher phrase**, but at the same time..

b. ...the rest of HP remains an **independent phrase** that behaves just like a phrase whose head has not moved — even though ...

c. The movement is **obligatory**. Movement of finite V to T in French satisfies some *need* of an element in this structure. The fact that the movement of the verb to T in French is triggered by properties of T itself can be seen from the fact that an infinitival verb does not move in this fashion:

No head-movement of infinitival V to T in French

(i) ne pas [acheter le livre]
prt not buy-INF the book
'to not buy the book'

(ii) *n'acheter pas [__ le livre]

The result is a word order in which negation precedes the verb just as it does in English *Mary will not read the book* or the bracketed clause of *I saw* [*Mary not buy the book*], which also lack V movement to T. Thus, the T-V relationship involves some **alteration in the features of T**.[32]

d. The zero-level head that undergoes head movement to another zero-level head ends up **tightly coupled** to its new host. The two heads end up behaving like a single morphologically complex word for later processes of grammar (both syntactic and phonological).

In French, for example, head movement of V to T has been claimed to account for the attachment of tense and agreement morphology to the finite verb (Emonds 1978, Pollock 1989). Once T has been attached to V by head movement, it is never separated from it by other processes (including further movement).

There is now an obvious question to ask about music. If PR is the musical counterpart to linguistic syntax, and is generated by Merge, does the formation of PR in music (like syntactic structures for language) involve Internal Merge in addition to External Merge? In what follows, we argue that at least head movement is indeed found in music, raising the possibility that *Merge in its full generality* is as characteristic of music as it is of language.[33]

---

[32] Although not directly relevant to the current discussion, we include a few words about how this approach works. An Agree relation (Chomsky 2000, 2001) between an unvalued tense feature of T and its valued counterpart on V assigns a value to the feature on T. As a language-particular consequence of this instance of Agree (*EPP*), V must Internally Merge with the T that it has just agreed with (Pesetsky & Torrego 2007).

[33] See Pesetsky (in prep.) for arguments in favor of *phrasal* movement in musical syntax. He proposes a more complex role for PR in which phrasal movement of harmonic constituents is the means by which musical rhythm is assigned to pitches.

5.2 Internal Merge in Music

5.2.1 Cadences

If there are instances of head movement in music, they should be identifiable by the same criteria that identify head movement in language, given the Identity Thesis. Specifically, we would expect to find musical phenomena with the following properties:

(42) **Properties that would diagnose PR head movement in music**

a. Some chord X must be performed **string-adjacent** to a chord Y. But at the same time...

b. ...X has a **normal set of syntactic dependents** of its own, linearized normally — and thus apparently also **heads its own phrase** (an XP).

c. The movement should be **obligatory**, insofar as it produces an **alteration in the features of Y** that is required in order for the derivation to succeed.

d. Even though X may take a normal set of syntactic dependents, **X is tightly coupled to its host Y**, such that they function as an indivisible unit for other purposes (cf. the notion *word*).

In this section we introduce a phenomenon from music that we claim has exactly these properties: the ***cadence* —** in particular, the variety of cadence known as the ***full cadence***.

Cadences are sequences of chords in tonal music that play a special role in the structural organization of pieces. Cadences function as structural signposts that mark the *right boundary* of structural units such as phrases within a melody, and larger sections within a piece. Standard music theory recognizes a variety of types of cadence, classified by harmonic, rhythmic and melodic criteria. The so-called ***full cadence*** is a cadence that terminates with two chords whose roots (in the normal case) descend by the interval of a fifth **(**five scale-steps).[34] This is the type of cadence that we focus on here (and we will often use the simple term "cadence" to refer just to this type). Full cadences mark the right boundaries of especially important structural units (such as the exposition section of a classical sonata-allegro movement), and the ending of a tonal piece is almost always signaled by a full cadence (in some styles, by a *perfect authentic cadence*, a subtype of full cadence).

Though it is traditional to focus on descent by a fifth as the core property of a full cadence, such cadences are actually characterized by a *cluster* of seemingly unrelated properties. Clusterings of seemingly unrelated properties provide the key testing ground for models of any complex cognitive domain. Does the model under discussion suggest an *explanation* for why these properties cluster as they do — as opposed to a myriad of imaginable alternatives? The model of tonal music that we have been developing here is a variant of the proposals in GTTM — but crucially embedded within the Identity Thesis. Consequently, the question we must ask whether the clustering of properties found in the full cadence, when structurally represented as GTTM suggests, can be explained as the musical counterpart of similar clusterings found in syntax of language. If the answer is positive, then the perspective advanced in this article is strongly supported by the phenomenon in question.

---

[34] At least one other possible relation between the two chords appears to yield the sensation of a full cadence, as Dmitri Tymoczko (personal communication) points out: a descent by the interval of a *fourth* (four scale steps), where the first chord is minor. This possibility appears in Bach and is also common in popular music of the twentieth century and beyond.

In (43), we list the properties of a Full Cadence:

(43) **Properties of Full Cadence**

A ***full cadence*** is a sequence of two chords[35] (δ, τ) drawn from the same pitch-collection (scale), where δ (the ***dominant***) is a major triad whose root is higher by the interval of a fifth (= five scale steps) than τ (the ***tonic***) — with the following additional properties:

a. δ must be **string-adjacent** to τ, but at the same time...

b. ...δ may have a **normal set of syntactic dependents**, linearized normally. Thus δ also **heads its own phrase** (δP).

The presence of at least one dependent of δ is exceedingly common, and is called a **subdominant,** which we will indicate with **υ** (GTTM, pp. 192 ff.).[36] It is typically built on the fourth or second step of the scale, but other possibilities are also found.

c. The participation of τ in a cadence (δ, τ) is crucial to **establishing the key as τ**.

d. In PR, δ behaves as if it were structurally directly subordinate to τ, even when its level of *prosodic* prominence (i.e. its RD number in TSR) should motivate a lower PR attachment site. GTTM calls this property **cadential retention.**

As is obvious, the properties of the full cadence listed in (43a-b) correspond precisely to the first two diagnostics of head movement stated in (41a-b) for language, and (42a-b) for music. We will suggest not only that this similarity is significant, but also that the remaining two properties of the cadence listed in (43c-d) instantiate the remaining two properties of head movement. In particular, we will suggest that the key-establishing property of the cadence stated in (43c) is the musical counterpart of the feature-altering process that has been claimed to be a prerequisite for Internal Merge in language; and that "cadential retention", as described in (43d), is just another name for the "tight coupling" characteristic of the output of head movement in language.

Finally, in section 6, we will compare our treatment of cadences to GTTM's. Reflecting the centrality of the full cadence to the harmonic flow of Western tonal pieces (or at least those from the "common practice" period), the GTTM model ensures that both δ and τ, when present, are marked as prominent in TSR. As discussed in the previous section, TSR prominence not only captures intuitions about rhythmic and harmonic stability, but also roughly entails PR importance. In keeping with its "type-1" character, however, no element of the GTTM theory actually *requires* the presence of a cadence at any point in a piece. For related reasons, the theory does not specify the presence of a cadence as a necessary condition for the establishment of key (or any other property of musical

---

35 GTTM, p. 168 adds a third chord to its description of a cadence, by asserting that a cadence is always a cadence *of* something, which they call the "structural beginning" to the cadenced group. This notion will not figure into our discussion, hence our simplification here.

36 Note also that some of the chords that commonly precede δ are uncommon or impossible before τ (e.g. ii or I^6_4), which suggests that they must be structural dependents of δ. Intuitions about reduction reach the same conclusion, since the legitimate reductions of υ - δ - τ (e.g. IV - V - I) include δ-τ and τ, but not *υ-τ.

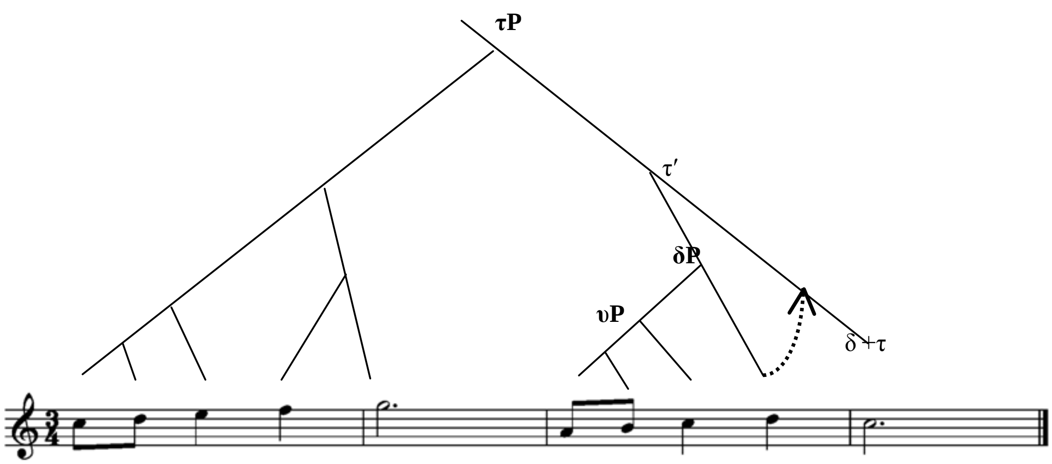
structure). Once we have sketched and supported our alternative treatment of cadences, we will be in a position to discuss and remedy these gaps in the theory.

5.2.2 Head movement and the adjacency restriction

We begin by sketching how head movement explains (43a-b). We assume that the TSR representation for a cadenced passage of music obeys the same general laws as any other musical passage, with the RD number for each musical event reflecting its relative prominence — as determined by meter, grouping and harmonic stability. We assume in particular that the TSR position of δ is determined solely by these general laws. In PR, the initial position of δ (i.e. the position in which it is Externally Merged) also reflects the same laws of musical structure that apply elsewhere. In particular, the initial position of δ will reflect the role of the Region Condition in minimizing disparities between TSR and PR constituency.

The special properties of the full cadence listed in (43) thus arise neither from the TSR position of δ, nor from its initial structural position in PR (the position in which it externally merges). Instead, these properties reflect the fact that after δ undergoes External Merge as part of the formation of PR, it merges a second time — undergoing Internal Merge with τ. This process obeys the same laws of head movement that are familiar from linguistic syntax. This is demonstrated in (44), with a toy melody devised for the purpose. The arrow indicates the Internal Merge of δ with τ. We use GTTM notation to indicate heads without committing ourselves to a theory of node labels in PR (a topic that we take up in the next section) — and offer below the tree an equivalent labeled-bracket representation of the relevant portions of the structure:

(44) **Cadential Internal Merge of δ to τ (Head-Movement)**



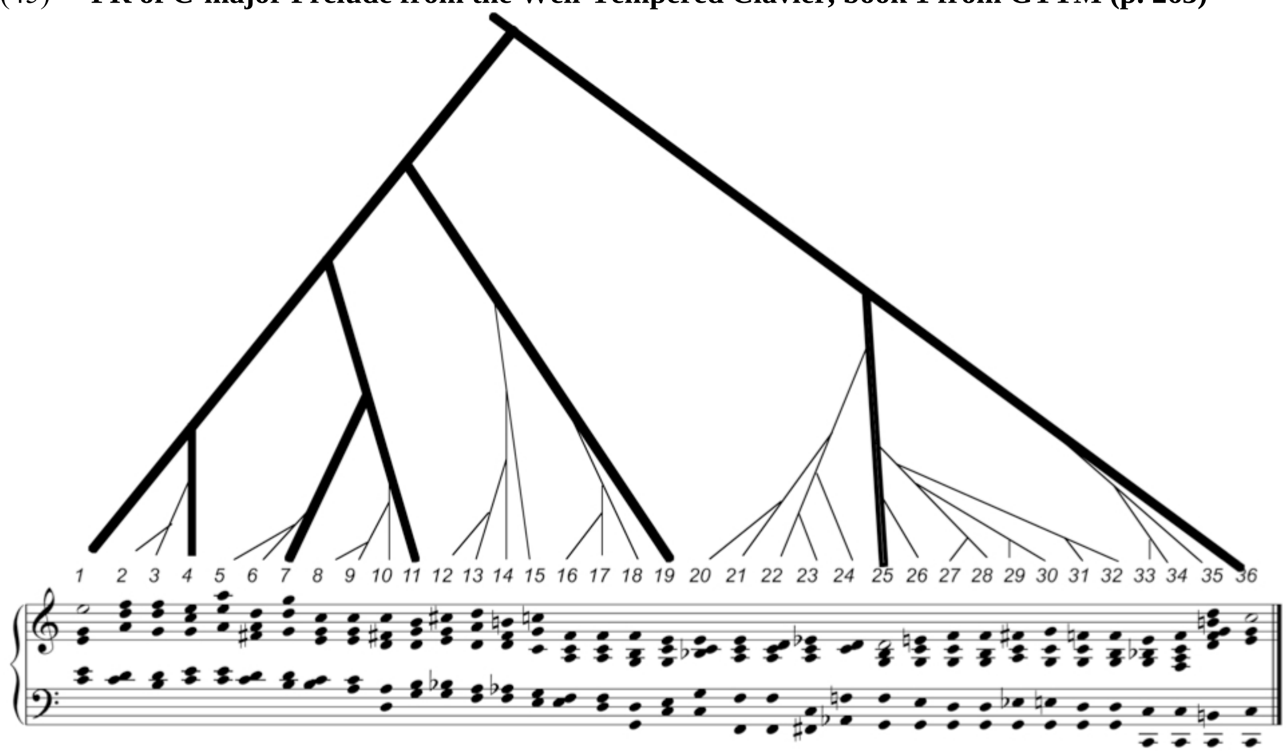
$[_{\tau P}$ … $[_{\tau'}$ $\mathbf{[_{\delta P}}$ $[_{\upsilon P}$ υ …] __ **]** $\mathbf{[_{\tau}\ \delta\ +\ \tau]}$

Crucially, as a consequence of the movement of the head δ to the head τ, the two chords end up string-adjacent. This property reflects the essence of head-to-head movement as seen in linguistic examples like (39), as analyzed in (40), in which the moving head merges directly with the target head. The claim that a full cadence involves head-movement of δ to τ thus entails strict string-adjacency between the two chords. Furthermore, although in its new position δ is not dominated by its own maximal projection δP, the structure of this maximal projection remains otherwise intact, and thus remains fully capable of hosting dependents such as the subdominant (υ).

It is important to emphasize how surprising and significant the string-adjacency condition on the full cadence is, from the perspective of GTTM and the Schenkerian tradition on which it builds. As discussed in section 4.3, in a structure built by (what we would now call) External Merge, the relationship between two constituents is determined by the heads of those constituents — whether or not the actual chords are string-adjacent. A constituent headed by an F-major chord that takes as its right-hand sister a C-major chord in the key of C will yield the same intuition of decreasing tension — no matter how many right-hand dependents of F or left-hand dependents of C might intervene linearly between the two heads. The hierarchical, headed nature of GTTM's structural representations correctly predicts the central discovery of Schenkerian analysis, the existence of "interactions at a distance". It is precisely because so much of tonal harmony allows for interaction at a distance that the string-adjacency requirement of the full cadence is so surprising.

To see the impact of the adjacency requirement, consider the PR structure proposed in GTTM for Bach's C-major Prelude from book 1 of the *Well-tempered Clavier*:

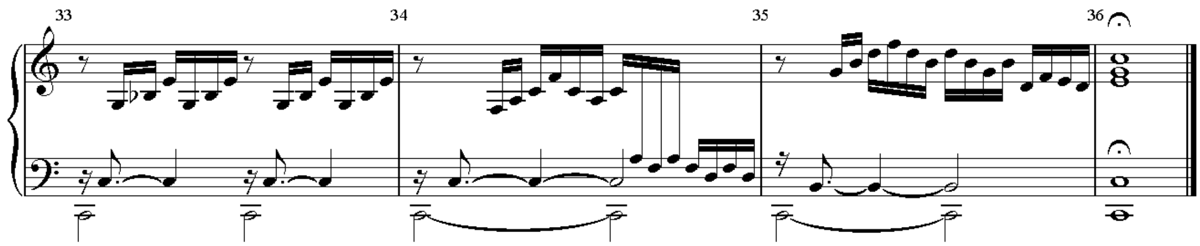
(45) **PR of C-major Prelude from the Well-Tempered Clavier, book 1 from GTTM (p. 263)**



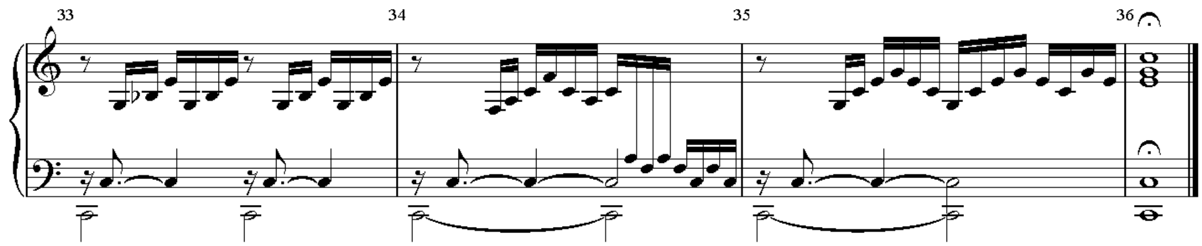
In this analysis, the dominant seventh G-chord in bar 25 (chord 24 in GTTM's numbering) heads a δP that is a syntactic left-sister to a projection of the final tonic C. In Lerdahl and Jackendoff's analysis of the Prelude (based closely on Schenker's (1932) analysis of the piece), this chord is identified as the "structural dominant" ("structural V"; p. 261) that plays the role of δ in the full cadence with which the piece concludes. In fact, however, an additional chord of the same sort (performed over a pedal-point C) is found in the penultimate bar (bar 35), where it immediately precedes the final tonic chord — and it is this penultimate chord, not the dominant in bar 25, that is crucial to the perception of piece as ending with a full cadence. Consider, for example, a variant of the prelude in which the dominant immediately preceding the final tonic is omitted or replaced by another chord, such as the tonic itself in (46b):

(46) **Ending of Bach C-major Prelude**

a. **original, with δ (bar 35) adjacent to final τ**



b. **rewritten without δ adjacent to final τ**[37]



Even if all the preceding bars of the Prelude remain unaltered, including the dominant chord in bar 25 whose maximal projection might still be parsed as a sister of the final tonic, the rewritten version is still not perceived as ending with a full cadence. The sense of closure and stability that a full cadence provides is missing, precisely because the final tonic is not immediately preceded by an instance of δ.[38]

This contrast between (46a) and (46b) is not specific to this piece, but exemplifies a general property of all tonal music (or at least the Western tradition of the past few centuries). If a chordal descent by five scale degrees is to be heard as a full cadence, it is never sufficient for a projection of the first chord to be merged as structural left-sister to the second. The head of the first chord must also immediately precede the head of the second chord — in the surface string, i.e. the musical surface. Interpose any other chord between a would-be δ and τ, and the result eliminates the sensation of a full cadence. As we noted above, an adjacency requirement of this sort is entirely surprising in a model like GTTM's that countenances only structures producible by External Merge. On the other hand, the

---

[37] The last beat of measure 34 has also been altered, to remove any hint of a dominant preceding the tonic in bar 35.

[38] Crucially, the alternative type of cadence discussed in note 34 (which descends by a fourth from a minor triad) is characterized by exactly the same cluster of properties (43a-d) as is the more typical cadence that descends by a fifth. Most strikingly, it too obeys the string-adjacency requirement. As Dmitri Tymoczko notes, one could rewrite the final bars of the C-major prelude to include the alternative cadence seen below:

If the penultimate chord is moved leftward so that is no longer adjacent to the final chord, however, the "perfect cadence" effect is lost.

existence of particular musical constructions in which two heads must be string-adjacent is unsurprising if musical structure, like linguistic structure, is derived by Internal as well as External Merge — in particular, the variety of Internal Merge known as head movement.[39] We return below to the question of why the full cadence in particular requires head movement of δ to τ, when we take up the final two properties of the full cadence, (43c-d).

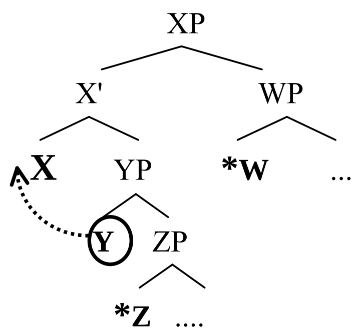
First, however, we should take note of one important difference between our argument from adjacency for head movement in music and the evidence for head movement in language. The adjacency requirement on δ and τ in the full cadence arises because δ is performed not where it was first externally merged, but in its final position — i.e. the position where it was internally merged. In this respect, the performance of δ follows the same pattern as the pronunciation of the moved phrase *which book* in (37)-(38) and the head-moved French verb in (39)-(40). In our linguistic examples, however, we had another source of evidence for head movement, and this source of evidence is not so clearly available in music. In our linguistic examples, we were able to pinpoint the position in which the moved elements would have been pronounced if they had not undergone movement. In the case of *which book* in (37)-(38), for example, we know that it would have been pronounced after the main verb, had it not moved; and in the case of the French verb, we know that it would have been pronounced immediately before the direct object. The fact that each of these elements is pronounced in an unexpected position, and that a "gap" is left where we expect them to be pronounced, raised the question to which Internal Merge proved to be the answer.

The fact that one can easily detect the original position as well as the final position of a moved element in language has also made it possible to discover a number of important constraints on internal merge in language. For example, it appears to be the case that whenever an element α moves to a constituent headed by β, β must c-command the initial position of α. Thus, in the case of *wh*-movement as in (37)-(38), the head C c-commands the initial position of the *wh*-phrase that merges with C'. Likewise, in the case of French verb-movement in (39)-(40), T c-commands the initial position of the moving verb. When c-command does not hold in this fashion, movement appears to be impossible in language. In the case of head movement, a tighter condition than mere c-command appears to hold across a wide variety of languages, the **Head Movement Constraint** (Travis 1984; see Baker 2001 for a non-technical presentation). This constraint limits head-movement of Y to X to configurations in which YP is the complement of X — i.e. a sister of the head X, as shown in (47):

---

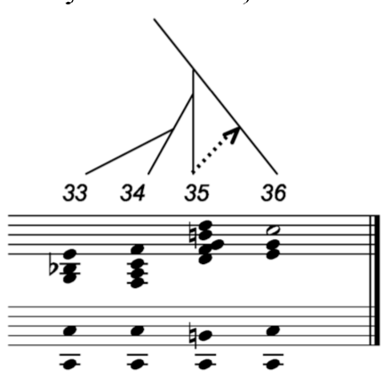
39 Our discussion omits one important deviation from this adjacency condition found in tonal music: the use of a chord built on the fourth degree of the scale (normally a IV^6_4) intervening in the musical surface between δ and τ (as in GTTM's (170) discussion of Bach's chorale *O Haupt voll Blut und Wunden* bars 5-6, fig. 7.15). Note, however, that when used in this context, this pitch event functions not as a normal IV chord, but as a cadential tonic with non-chordal tones in its inner voices (a suspension and a neighbor tone in this case). It is a structural *appoggiatura* — which arises in the musical surface without heading a PR constituent of its own. A number of peculiarities of the construction suggest that this is correct: 1. The chord IV^6_4 is not generally used in the Western tonal idiom except as a product of passing tones or appoggiatura notes. 2. The outer voices are both scale degree 1, as we would expect with the tonic of a perfect authentic cadence. 3. The inner voices are approached and resolve in the standard manner of non-chordal tones; a true IV chord could be followed by a variety of other material. 4. This chord can not license dependents of its own; for instance, it is never preceded by a V7 of IV. (See our discussion of tonicization and modulation below.). Thus, whatever licenses the appearance of this pitch event, it does not appear to count as a syntactic entity, even in the somewhat exceptional way that a V chord, which does license its own syntactic dependents, is a syntactic entity.

(47) **Effects of the Head Movement Constraint**
Y may undergo head movement to X in the configuration below. W and Z may not.[40]



As far as we can tell, the full cadence in music offers no counterexamples to either the general c-command condition on movement or Travis's tighter Head Movement Constraint. If, for example, we accept (45) as the basic PR structure for the C-major Prelude, with the addition of internal merge of chord 35 to the final tonic 36, this instance of internal merge obeys both constraints. Chord 36 c-commands chord 35, since 35 is dominated by all nodes that dominate 36. Furthermore, the maximal projection of 35 (dominating 33, 34 and 35) is the complement of 36, so δ-to-τ movement of 35 also obeys the Head Movement Constraint. The relevant part of the structure, including the head movement of chord 35 to 36, is diagrammed in (48):

(48) **δ-to-τ head movement in the final two bars of the C-major Prelude (bars 35-36)**[41]
(indicated by dotted arrow)



Notice now that if (48) correctly indicates not only the structure and final linearization of the terminals of the tree, but also how chord 35 *would* have been linearized if no head movement had taken place — then the movement seen here differs from the instances of movement that we examined in (37)-(38) and (39)-(40) in being *string-vacuous*. That is, in moving to chord 36, chord 35 does not cross over any linearly intervening material (unlike the *wh*-phrase in (37)-(38) and the French verb in

---

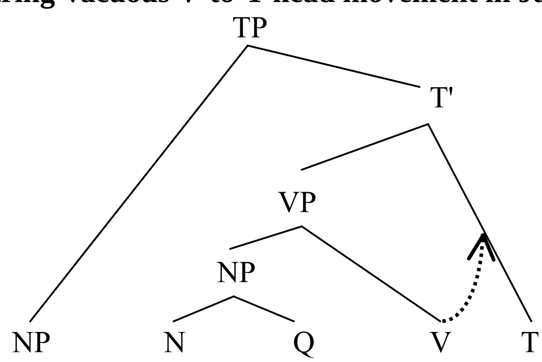
[40] Z may, however, undergo head movement to Y, and Y+Z together may head-move to X. This happens, for example, when a verb moves to T, and then the verb — with tense morphology affixed to it — move to C, yielding a "subject-verb inversion" construction familiar from questions and other constructions in many languages.

[41] We discuss bars 33 and 34, which also show δ-to-τ movement (as an instance of "tonicization"), omitted here.

(39)-(40)). One might worry that this difference might reflect an unexpected and unwelcome distinction between language and music, given the Identity Thesis. In fact, however, this is not the case.

Though it has occasionally been suggested that string-vacuous movement is prohibited in language (for example, by Chomsky 1986a, 48 ff.), string-vacuous movement does seem to be possible after all. In Japanese, for example, a "head-final" language in which the heads of phrases follow all their dependents (cf. chord 35 in (48)), Koizumi (1995, 165 ff.; 2000), following Otani and Whitman (1991), has argued at length that the verb moves to T, just as it does in languages like French, despite the fact that such movement does not cross other overt material:

(49) **String-vacuous V-to-T head movement in Japanese**



Mary-ga ringo-o 3-tu kat- -ta
Mary- NOM apple-ACC three-CLASS buy PAST
'Mary bought John three apples.'

Koizumi's evidence for the string-vacuous head movement seen in (49) is the ability of the elements of the verb phrase, crucially minus the verb itself, to behave as a constituent for a variety of syntactic tests (including coordination, cleft sentence formation, and several others).[42] Likewise when a subject is questioned in English (e.g. *Who read the book?*), there are reasons to believe that *wh*-movement has applied, and that the subject has internally merged in the same position as that seen in (37)-(38) — even though this instance of *wh*-movement does not cross over other material, and is thus string-vacuous:

(50) **String-vacuous *wh*-movement in English**

Who C [$_{TP}$ __ T [$_{VP}$ read the book]]?

One argument comes from observations by Pesetsky (1987, 125, note 20) and Ginzburg & Sag (2002, 237) concerning the distribution of "expressive *wh*-phrases" such as *who the hell* or *what on earth*.

---

[42] Alternative accounts of Koizumi's data (and related phenomena) that eschew string-vacuous head movement have been suggested by Takano (2002) and by Fukui & Sakai (2003). See Kishimoto (2008) for a survey of this discussion, which we will not address here. Kishimoto (2007) himself has argued that V does not raise to T in Japanese (i.e. that (49) is wrong) — but that negation (*na(i)*), a head structurally situated between V and T, does. (He also argues that this instance of head movement has detectable semantic effects as well.) If Kishimoto is correct, we should replace (49) with a comparable negative sentence, to illustrate the possibility of string-vacuous head movement. Our arguments concerning head movement in language and music would be unaffected by this substitution.

Phrases of this sort must undergo *wh*-movement, as can be seen by comparing (51a) with (51b). Crucially, a subject *wh*-phrase patterns with moved phrases as in (51a), and not with the unmoved phrase in (51b), as (51c) shows — even though the movement posited for the subject *wh*-phrase is string-vacuous:

(51) a. **expressive on moved non-subject *wh*-phrase**
[Who the hell] did you introduce __ to whom?
b. **expressive on unmoved *wh*-phrase**
*Who did you introduce __ to [who the hell]?
c. **expressive on subject *wh*-phrase**
[Who the hell] __ read the book?

We may therefore conclude that string-vacuous δ-to-τ movement in music, as in (46), is compatible with the Identity Thesis, since movement in language (including head movement) may apply string-vacuously as well.

On the other hand, despite the apparent compatibility of our proposal for δ-to-τ movement with what is known about head movement in language, we do face certain difficulties in putting our proposal to the test that corresponding linguistic investigations do not. In particular, difficulties arise when we attempt to establish the precise position that δ has moved from. In language, the differing and sometimes idiosyncratic properties of individual lexical items allow us to construct examples in which we know that a particular word or phrase would be heard in a particular position, were it not for movement — because some lexical item elsewhere in the structure requires it. An obligatorily transitive verb whose object is not pronounced in its expected position, for example, immediately raises the question of what happened to the object. If the missing object is pronounced elsewhere in the utterance, and turns out to be a *wh*-phrase, we can immediately guess that *wh*-movement has dislocated the phrase from its expected position, and moved it to the position in which it is actually heard. We can then test predictions of this hypothesis with reference to both positions. Likewise, when a French phrase is heard that sounds like a Verb Phrase — except that the verb itself is missing from the Verb Phrase, but heard in a VP-external position where we expect to hear tense — we can posit verb-to-T head movement, and once again test predictions of this hypothesis with reference to both positions.

Because music lacks a lexicon, it also lacks the cooccurrence restrictions found in language that allow us to localize not only the surface position, but also the underlying position of a moved element. That is why (taking our analysis of the full cadence as an example) we cannot truly pinpoint the position where we would have expected to hear δ if it had not undergone cadential head-movement. Consider, for example, a standard musical "IV-V-I" (υ - δ - τ) cadential progression, in the context of our proposal that full cadences involve δ-to-τ movement. Assuming that IV and V form a constituent from which V head-moves to I, there is no independent property of IV or V (and no independent syntactic process sensitive to other properties of the IV-V constituent) that can tell us, for example, whether, had V not moved, it would have been the left-daughter or the right-daughter of the IV-V constituent. Likewise, in examples like the ending of the C-major Prelude, δ-to-τ movement might be string-vacuous as discussed above, with the crucial δ merging first in exactly the position shown in (45) — but we cannot be sure, since we do not have an independent means of verifying the position of δ before movement. Musical δ-to-τ movement thus contrasts with linguistic counterparts such as French finite V-to-T movement, where the position where the verb would have been pronounced if movement had not occurred can be determined quite precisely, on the basis of our independent knowledge of the selectional properties of the other lexical items in the structure.[43]

These facts do not logically weaken the argument from adjacency for δ-to-τ head movement, but they do mean that this argument cannot be bolstered by some of the additional considerations that are available in the investigation of language.[44] For now, therefore, we must rely on other types of evidence to strengthen the case for δ-to-τ movement beyond the argument from adjacency — to which we now turn.

5.2.3 Key establishment and feature alteration

Syntactic movement in language appears to be driven by the needs of one or both of the constituents involved in the movement operation. As we noted in (41c), for example, the movement of the verb to T in French appears to be triggered by properties of T, since the verb only raises when T is finite (i.e. bears tense) — and raises obligatorily in such cases.[45]

Similar observations can be made about many — perhaps all — types of syntactic movement. *Wh*-movement, for example, appears to satisfy requirements of C, rather than a requirement of the *wh*-phrase that moves. In an English "multiple question" that involves more than one *wh*-phrase, *wh*-movement must take place, but only one of the *wh*-phrases undergoes this movement (typically the structurally highest phrase, a much-discussed property of such constructions). Example (51a) already displayed a construction of this sort, as does (52). In example (52a), we can see that failure to move

---

[43] Comparative investigation of other musical idioms could in principle help. If an idiom were similar to Western tonal music, except without head movement in full cadences, we could observe where the crucial δ is found in that idiom, and use these conclusions to enrich our knowledge of the underlying structure of cadences in idioms that do require head movement. Certain blues idioms in which V-IV-I forms a standard ending might in fact constitute such an example (and might ultimately yield an argument that the δ-to-τ movement found in non-blues tonal idioms is not necessarily string-vacuous). Since the overall tonal space of these idioms, however, is different from Western tonal music (for example, it is pentatonic and modal), we must remain cautious for now about drawing conclusions.

[44] It is perhaps surprising to note that our GTTM-based Schenkerian arguments for External Merge in music come from the existence of long-distance interactions among musical events (heads of sisters separated by non-head dependents), while our argument for Internal Merge comes from the existence of a strict adjacency requirement between two musical events. The norm in elementary presentations of linguistic syntax is just the opposite. Nonetheless, the existence of long-distance head-to-head interactions is predicted by External Merge if it produces headed structures, as we have supposed — and in the absence of clear evidence concerning the initial position of δ in δ-to-τ cadential movement configurations, it is the adjacency condition on the output that offers the best evidence that a different type of Merge has taken place. Precisely because music and language differ in their building blocks, it should come as no surprise that the nature of the most available evidence for phenomena common to the two domains may differ too.

[45] In highly elevated registers of French, the verbs *être* 'be' and *avoir* 'have' may raise to infinitival T as well, as noted by Pollock (1989, 386-391). In this respect, the infinitival T of this register of French resembles the finite T of English, which attracts an auxiliary verb if one is present in the structure, but otherwise does not attract anything. In some ways, the behavior of a tonic chord more resembles this variety of T than it resembles French finite T, since a tonic chord does not need to form part of a perfect cadence. We will not develop this point further here.

any *wh*-phrase yields an unacceptable result; example (52b) is acceptable because one *wh*-phrase has moved; and example (52c) is unacceptable because more than one *wh*-phrase has moved.[46] These constructions must involve an interrogative flavor of C that imposes these requirements, because the main-clause verb *wonder* selects for a complement whose head is this C:

(52) a. **No *wh*-movement**
*Mary wondered [$C_{+interrog}$ Sue gave what to whom].
b. **One instance of *wh*-movement**
Mary wondered [what $C_{+interrog}$ Sue gave ___ to whom].
c. **More than one instance of *wh*-movement**
*Mary wondered [what to whom $C_{+interrog}$ Sue gave ___].

A non-interrogative C, on the other hand, does not allow a *wh* phrase to internally merge with its projection.

To put these observations slightly differently: the input to movement in each case includes a higher head H, with certain featural needs. The output of movement contains the same head H with those needs satisfied. As a consequence, H will "look different" after movement, and may be treated differently by subsequent syntactic processes. It has been argued, for example, that verb movement to T in French satisfies a requirement of tense morphology (such the future *-r-* in (40)) that it be suffixed to a verb-stem. We may imagine that as a consequence of verb-movement, a previously unvalued [*suffixed*] feature on T comes to be marked [+*suffixed*] — satisfying a general requirement of the grammar that such features be positively valued.

Returning to our discussion of head movement in the full cadence, we must now ask whether an instance of τ to which δ has moved also displays properties after movement that it would not otherwise come to display. If we find evidence that δ-to-τ movement changes the character of τ, as V-to-T movement alters the features of T and *wh*-movement alters the features of C, we will have discovered another deep property that the full cadence shares with head movement in language.

We have already noted informally the role of the full cadence in establishing a key. We propose that this property of the full cadence arises from the valuation of a feature on τ in a manner precisely analogous to the linguistic phenomena just discussed. In particular, we suggest that cadential δ-to-τ movement has the function of **tonic-marking** τ, i.e. assigning it the feature [+TON] . When a head τ in a structure K is tonic-marked by δ, it has the consequence of allowing the terminal nodes of a particular subtree of K (determined by τ) to be understood as belonging to *the key of τ*.[47]

---

[46] The Slavic languages, among others, allow (and in some cases, require) the pattern seen in (52c). An extensive literature is devoted to determining the ways in which the demands placed by C on accompanying syntactic structure are met in different languages, so as to explain these patterns of variation. See, for example, Richards (1997, 2001) and Pesetsky (2000) for discussion.

[47] The Western tonal tradition has developed a number of conventions surrounding tonic-marking that serve to signal an upcoming cadence. (These include the use of otherwise unusual chords and a restricted set of voice-leading options.) It is also not uncomment for a composer to exploit the strong expectations created by these conventions for artistic effect — as when a well-prepared potential δ is not followed by the expected τ, as in the so-called "deceptive cadence" (which is not in fact a cadence at all, in our terms). We are grateful to Ray Jackendoff for raising the question of the "deceptive cadence" in this context.

The notion of "key" is central to the Western tonal idiom. The choice of key for a given passage determines, first and foremost, the set of pitches used to build the passage — that is, the **pitch collection** (= **scale**) from which its terminal nodes are drawn[48]. The pitch collection of a passage in turn determines the abstract **pitch space** within which its notes and harmonies are situated — a multidimensional ranking of the pitches and harmonies built from them that governs perceptions of relative tension and stability in ways that are crucial to such matters as the choice of head for PR and TSR constituents. These matters are discussed in considerable detail by Lerdahl in TPS, and we assume his proposals here.

Let us now consider the relation of tonic-marking to key assignment in further detail. If a head bears the feature [+TON], every projection of that head will also bear the feature [+TON] (by the principles discussed in section 4.4). It is therefore natural to think of a node bearing the feature [+TON] as establishing a "key domain" that assigns the elements that it dominates to a particular key. A first approximation is given in (53) and (54):

(53) **Key Domain (version 1)**
A node marked [+TON] is a key-domain.

(54) **Key Establishment (version 1)**
If α is a key domain that dominates *x*, *x* is assigned to the key of α

In simple cases, this proposal yields the correct result. Rule (54), for example, will assign every element of the toy melody (44) to the key of C, as a consequence of δ-to-τ head movement that creates a full C-major cadence. As predicted, every pitch event in (44) is drawn from the C-major pitch collection {C, D, E, F, G, A, B}.

Rule (54) yields an incorrect result, however, in cases of ***key-domain recursion***, where one key domain is embedded within another, as in (55), where β and α are distinct tonic-marked elements:

(55) **Key domain recursion**
$[_{\alpha,\ +\mathrm{TON}}$ ... X ... $[_{\beta,\ +\mathrm{TON}}$ ... Y ...] ... Z ...]

In a structure like (55), rule (54) will incorrectly assign Y simultaneously to the key of α and to the key of β, and incorrectly predicts that Y must be a member of both the α pitch collection and the β pitch collection. In fact, a listener will hear a passage with the structure of (55) as beginning in the key of α, moving into the key of β, and then returning to α; — and Y will belong to the pitch collection of β, not α. This is in fact the well-known and intensively studied phenomenon of ***tonicization*** or ***modulation*** (a temporary shift in key). We thus revise (54) to (57):

(56) **Key Establishment (final version)**
If *α* is the *smallest* key domain that dominates α*, x* is assigned to the key of α.

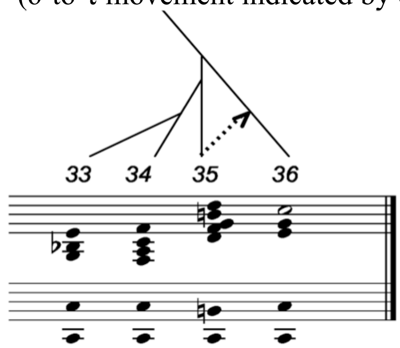
A simple example can be seen in final four bars of the C-major Prelude already displayed in (46) (and at many other points in the Prelude as well). Because the penultimate chord of the piece (chord 35)

---

48 "Non-chordal" phenomena such as chromatic passing tones may constitute exceptions to this rule, but because they are highly restricted in use and function in standard tonal idioms, we will not discuss them here. Pesetsky (in prep.) does discuss related phenomena in the context of the overall modal developed here.

undergoes δ-to-τ movement, tonic-marking the final C-major chord, and because the final chord is the head of entire composition, the piece as a whole is in the key of C-major. Note now, however, that chord 33 contains the pitch B♭, which does not belong to the C-major pitch collection — but does belong to the pitch collection of F, the root of the chord immediately to its right (chord 34). As it happens, the root of chord 33, C, is also five scale degrees higher than F,[49] so if chord 33 undergoes head movement to chord 34, it can tonic-mark it. (In traditional terminology, it functions as a **secondary (**or **applied**) **dominant** of F**.)** After δ-to-τ head movement of chord 33, the smallest key domain that contains chord 33 (in both its pre-movement and post-movement locations) is F rather than C, so the presence in the chord of B♭ from the key of F (rather than B♮, from the key of C) is no longer surprising:[50]

(57) **Tonicization in bars 33-34 of the C-major Prelude**
(δ-to-τ movement indicated by dotted arrows)



A much more extended example of the same phenomenon can be seen in bars 5-11 of the Prelude. These bars, in GTTM's analysis, form a constituent headed by the maximal projection of the G-major chord in bar 11. In this passage, δ-to-τ movement of chord 10 to chord 11 tonic-marks the latter. As a consequence of rule (57), bars 5-11 are therefore assigned to the key of G. This accords with our intuitions about the passage (that the music modulates from the key of C to the key of G — and then modulates back), and explains the presence of F♯ from the G-major pitch collection (rather than F♮ from C-major) in the bars under discussion.

Our proposal still makes one further undesirable prediction. This problem arises because, as formulated, (53) treats every [+TON] node as a key-domain, which appears to be too strong.[51] Consider the standard cadential formula IV-V-I discussed in (43b), in which a subdominant (υ) is a left-dependent (as argued in GTTM) of the δ that head-moves to τ, as diagrammed in (58a) below. A common expansion of this formula adds a secondary dominant (V/V or V7/V) between the main subdominant and the main dominant, as shown in (58b), where dotted lines once again indicate δ-to-τ head movement:

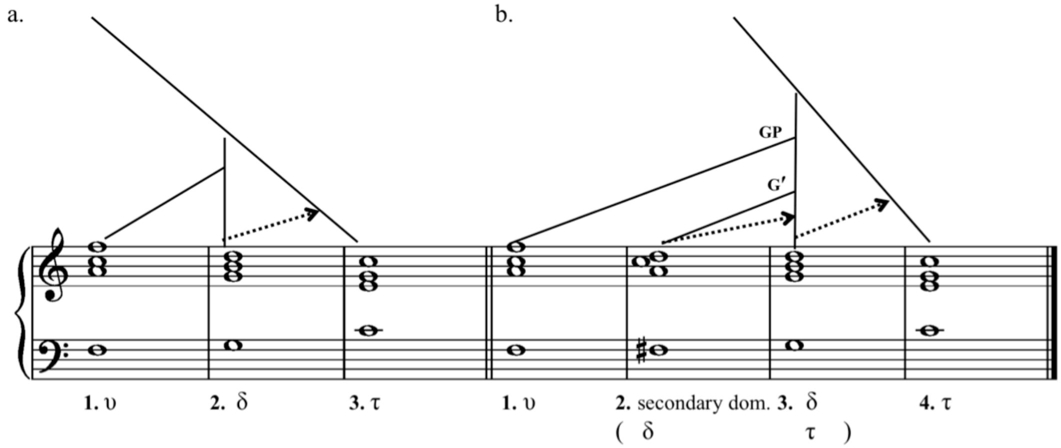
---

49 We make the traditional distinction between the *root* of a chord and its lowest note (the *bass*).

50 The term "full cadence" is unlikely to be used to describe 33-34, because no structural section boundary is demarcated by 34. The term is normally reserved for passages that both share the formal properties we have attributed to the cadence and the function that it may easily assume of marking the right edge of a structural boundary. This is a purely terminological detail, but it does make clear that the form and the function of the cadence are related, but logically separable properties.

51 We thank John Halle, who brought this issue to our attention.

(58) **IV-V-I cadential formula and extended version with a secondary dominant**



If every [+TON] node is a key domain, then (54) will not only correctly assign the secondary dominant seventh chord in bar 2 of (58b) to the key of G (i.e. identify it as V^{6}_{5}/V) , it will also *incorrectly* assign chord 1 of (58b) to the key of G as well — since it is dominated by the maximal projection of the tonic-marked chord 3 (GP). We know that this is an incorrect consequence for at least two reasons. First, chord 1 contains F♮, which belongs to the C-major pitch collection but not to the G-major pitch collection (which contains F# instead; cf. chord 2). Second, as discussed above, the cadential formula in (58b) is identical to (58a) except for chord 2, and is perceived as an extended version of the same formula. To assign chord 1 to different keys in the two passages (even if the choice of pitches had permitted this) would fail to capture the intuition that adding the secondary dominant leaves the nature and function of the other chords unchanged.

These observations indicate that it is not obligatory for a tonic-marked node to act as a key-domain. In (58b), for example, the non-maximal projection G' acts as a key-domain, assigning chord 2 to the key of G, but the maximal projection GP does not. Consequently, we revise (53) to ensure that key-domain status for [+TON] nodes is optional, rather than obligatory:

(59) **Key Domain (final version)**
Optional: A node marked [+TON] is a key-domain.[52]

Finally, consider the question of which key chord 3 is assigned to in (58b). Since G' is a key-domain, chord 3 must be assigned to the key of G. This is not in and of itself an incorrect conclusion, since it is tonicized by the preceding chord. At the same time, however, it clearly also fulfills the same function in (58b) that its counterpart fulfills in (58b), and thus also belongs to the home key of C. In fact, our proposal as stated already captures this dual nature of tonicized chords like chord 3 in (58b). Although *chord 3* will be assigned to G-major by virtue of domination by the key domain G', neither G'

---

52 Note that the fact that δ should be a major triad does not follow from cadential [V-I] being required to be interpreted as a Key Domain, since this is optional. It must follow instead from an independent requirement. That such a requirement exists is shown by the fact that it holds of [V-I] cadences in minor keys as well (and may also be a requirement satisfied by VII-i modal cadences in other idioms). It's possible that the process of Tonic-marking affects δ in addition to τ, with the raised third of δ being the result. Alternatively, a requirement for leading tone might outrank faithfulness to pitch-collection. Note that the opposite constraint seems to hold of δ in the [IV-I] cadence discussed in footnotes 34 and 38: it must be a minor triad. We leave this as an unsolved puzzle.

itself (because the domination relation is not reflexive) nor GP will be assigned to G-major. Instead, they will both be assigned to the key of C, since they are dominated by projections of the final tonic. Though the chord itself (in its original unmoved position) is a tonic of G-major, the maximal projection that it heads functions as a dominant of C, exactly as required.

We have thus seen ample evidence that δ-to-τ movement is associated with an alteration in the features borne by the target of movement (the chord τ) — just as is the case in language. We have supported this claim by exploring in some detail the precise way in which a tonic-marked chord affects other aspects of PR structure, such as the pitch-collection from which its building blocks are chosen. By showing that a coherent proposal concerning key assignment can be built on the idea that δ-to-τ movement marks τ [+TON], the case for our analysis of the full cadence as an instance of head movement is correspondingly strengthened. We may thus add a third property of the cadence, (43c), to the list of ways in which the properties of the full cadence reflect the properties of head movement already familiar from research on language.

5.2.4 Tight coupling of δ to τ

GTTM devotes considerable attention to a problem that is specifically raised by cadences for the mapping between TSR and PR. The harmonic importance of a cadential element may be greater than its rhythmic importance would otherwise warrant. This is particularly true of δ in the full cadence, which often occupies a position that would otherwise be treated as deeply embedded with respect to the following τ — due to such factors as δ's metrical weakness or parallelism with deeply embedded elements in other groups (often contrasting with the high prominence of the following τ).

In response to this problem, GTTM outlines a special procedure that promotes δ to a more prominent position in TSR than more generally applicable principles of prosody would license — with consequences for PR due to the heavy constraints on the mapping between these two components. Under GTTM's proposal, when two chords δ and τ comprise a cadence, and τ would otherwise be the head of a time-span S, a special procedure allows δ and τ together to function as ***joint heads*** of every projection of S except the highest.[53] This phenomenon is known as ***Cadential Retention,*** indicated in GTTM with a hollow oval connecting δ and τ in TSR, where relevant. Above the domain of Cadential Retention, τ projects alone. PR does not treat δ and τ as coupled, but for the purposes of the Region Condition, the highest point at which δ and τ were coupled in TSR determines RD(δ).

Cadential Retention is motivated in GTTM by the observation that the two chords of a cadence function as a unit. In particular, if the τ of a cadence heads a time-span, the intuitions of prominence that adhere to τ also adhere to δ. For example, in performing reductions of a passage that terminates in a full cadence, δ will be the last chord to disappear before τ itself.[54]

---

[53] GTTM's well-formedness rule TSR-WFR4 (p. 159) actually entails that δ and τ function as co-heads of *all* projections of S, but the diagrams included in GTTM show the co-head status of δ ending one projection-level lower than the rule suggests— with the consequence that what GTTM calls the "structural beginning" of the "cadenced group" in TSR takes a bare τ as its sister, rather than δ and τ functioning as co-heads. We will assume here that the rule rather than the diagram is correct.

[54] The fact that the middle member of Schenker's *Urlinie* is always an instance of δ (flanked by tonics) reflects this intuition as well — though in some cases (including GTTM's own Schenker-derived TSR structure for the C-major Prelude, p. 262) the proposed source of this remnant dominant will not be the adjacent dominant that we would regard as the true δ of the full cadence.

The structural limit on cadential retention reflects the fact that past the TSR domain headed by the entire cadence, only τ seems to project. Only τ, and not δ, is relevant to the stability conditions that pick TSR heads at higher levels of structure.

Finally, the fact that the coupling of δ and τ is not visible to PR reflects the observation that δ takes its own syntactic dependents (instances of υ) that are not also dependents of τ. The fact that the TSR coupling of δ and τ is not visible to PR also captures the fact (mirrored in TSR) that the properties of a constituent headed by cadential τ are determined by τ alone (and not at all by δ).

The output of Cadential Retention in GTTM's theory of TSR bears an obvious resemblance to the head movement of δ to τ that we have posited in PR. If we suppose that the output of δ-to-τ movement, like the output of head movement in language, produces a unit whose elements are not just adjacent, but "tightly coupled" like the morphemes of a word, we expect this to be reflected in TSR. TSR should retain the tight coupling produced by head movement in PR, just as the prosodic structure of language retains the tight coupling among the morphemes of a morphologically complex word. Linguistic prosody does not, for example, place major prosodic boundaries between the morphemes of a word (at least in 'normal' discourse contexts), but treats the morphemes of a word as an indissoluble unit. In much the same way, we expect TSR to treat the link between δ and τ as indissoluble.

At the same time, there is a conceptual difference between GTTM's Cadential Retention and the tight coupling that stems from our proposed δ-to-τ movement that favors our approach over GTTM's. GTTM's proposals are *sui generis*, invoking mechanisms specific to musical cadences — and nothing else. By contrast, our proposal extends to music a syntax independently observable in a variety of linguistic phenomena.

GTTM's approach, furthermore, makes an apparently false prediction about PR that our approach avoids. By directly elevating the prominence of δ in TSR, the GTTM theory predicts that in PR, δ will choose its sister from among the other elements with extremely high TSR prominence. (Recall our discussion of GTTM's TSR-PR mapping in section 4.6.) Consequently, δ should be able to supply the right boundary for extremely large prolongational regions — regions potentially almost as a large as those supplied by the final τ (which, as discussed below, usually takes as its most prominent dependent the initial τ of a piece) In a passage of any length, the left-dependent of δ (for example, a IV functioning as υ) would be expected to be quite remote from δ as a matter or course, just like the left-dependents of cadential τ. As far as we can tell, this situation seldom, if ever, arises (as an inspection of the analyses of specific compositions included in GTTM, for example, suggests.)[55] By contrast, no such prediction is made in our proposal. In PR, δ is actually attached quite low (as the lowest sister of τ), and by virtue of the Region Condition will be quite low in TSR as well. Consequently, its syntactic dependents are expected to be fairly local, as is indeed the case.

If it is a tight coupling of δ to τ that accounts for the intuitions that motivated Cadential Retention in GTTM, and if this tight coupling is a direct consequence of δ-to-τ movement in PR, as we have argued, then all four properties of head movement in language listed in (41) have now been discovered in the musical syntax of the full cadence. If this is so, then Internal as well as External Merge plays a role in building musical structures, and thus supporting the Identity Thesis.

---

[55] GTTM's stipulated preference for "normative prolongational structure" (discussed below) is the main factor that prevents this from happening, favoring attachment of material leftwards to the initial τ (called a "departure") over rightward attachment to δ, in situations where both choices are in principle available.

6 The Tonal-Harmonic Component

6.1 Key domains and Full Interpretation

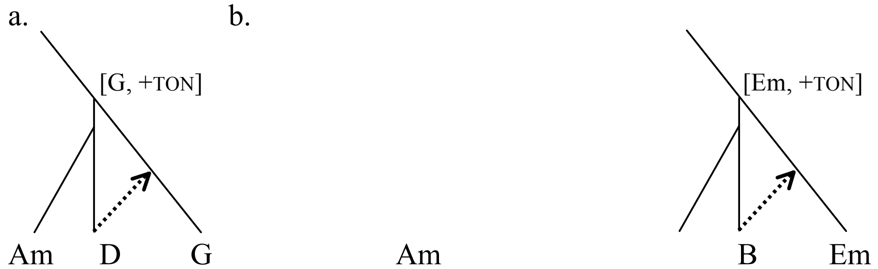
In the previous section, we proposed the optional rule (59) that identifies a constituent as a *key domain*. What kind of a rule is (59) , and what component of the grammar does it belong to? It cannot be a rule that assigns a feature to a node, the way tonic-marking assigns the feature [+TON], or the way pitch-class (e.g. 'E♭') functions as a feature. All projections of a given head share all features. The kind of fact exemplified by (58b) already made it clear that the property of being a key domain is not a feature in this sense, since the fact that the intermediate projection of chord 3 constituted a key domain did not entail that its maximal projection had the same status. Indeed, the two projections belonged to distinct keys.

The selection of a particular tonic-marked node α as a key domain does not in fact affect α at all. Instead, it provides a framework for the harmonic interpretation of other nodes — the nodes dominated by α. The natural way to understand (59), therefore, is as a rule of *interpretation* by a harmonic system separate from the syntactic system proper.

Suppose, for instance, a piece contains an a-minor chord. The syntactic system that we have been discussing is responsible for allowing this chord to undergo Merge with other chords, and for identifying it or one of its sisters as the head of the constituents produced by Merge. The interpretative system that includes rule (59) has a different job. This system allows for the identification of the chord in terms of scale degree within a given key. For example, it is this interpretive system that decides whether an a-minor chord in a given context is a submediant in C-major, a supertonic in G-major, or a subdominant in e-minor. It is this system as well that ensures that an A-minor chord in a C-major context has the same harmonic properties as a D-minor chord would have in an F-major context. Informally speaking, this component is the locus of "Roman numeral notation" (which individuates chords according to the scale-degree of their root in the current key) — and of all harmonic factors that depend on Roman numerals.

For instance, this harmonic component will identify the a-minor in (60a) as a supertonic in G-major, where it is dominated by a [+TON] G node. It will identify the a-minor in (60b) as a subdominant in e-minor, where it is dominated by a [+TON] e-minor node.

(60) a. b.



Let us call the component that includes the Key Establishment rule in (56) and the Key Domain rule (59) the **Tonal-Harmonic Component (THC)**. If we are correct, the THC stands in a relation to PR much like that explored in section 4 between TSR and PR. We suggested there that TSR stands to PR in music as prosodic structure stands to syntactic structure in language. In the spirit of that discussion, we now explore the possibility that the status of THC in music is analogous to that of *semantics* in language.[56] Recall from the boxology of (7) that linguistic semantics is an interpretive component, and that the process of semantic interpretation makes reference to details of syntactic structure — much as the Key Domain Rule and the notion of Key Membership make reference to the structures of PR.

In section 5.2.1, we noted two problems with GTTM's treatment of the notion *cadence* that we related to the distinction between type-1 and type-2 theories with which we began this paper. First, we noted that in keeping with its "type-1" character, the GTTM theory does not *require* the presence of a cadence at any point in a piece. Second, we noted that the theory "does not specify the presence of a cadence as a necessary condition for the establishment of key (or any other property of musical structure)". If we are correct, viewing the Key Establishment rule (56) and the Key Domain rule (59) as rules of THC takes care of the second of these problems. Let us now turn to the first.

If the presence of a cadence is a precondition to tonic-marking, and tonic-marking a precondition to establishing a Key Domain, and if the presence of a Key Domain is crucial to the notion of being "in a key" — then the observation that every piece must be in some key can be seen as a side-effect of the more general fundamental principle in (61), coupled with the claim in (62) (and the fact that pitch-classes are the building blocks of music):

(61) **Principle of Full Interpretation (FI)**
Every constituent must receive an interpretation in THC.

(62) **"Interpretation" (in Western tonal music)**
A pitch event is interpretable only in relation to some local tonic.

Strikingly, the principle in (61) has in fact been argued to hold in language as well, with the substitution of "semantic component" for THC (Chomsky 1986b) and with a semantically based notion of "interpretation" replacing (62). It is the Principle of Full Interpretation for language that dictates that a meaningful morpheme like *cat* cannot be merged in the syntax unless its semantic value is used by the semantics — and it is for the same reason that a reflexive pronoun, which comes from the lexicon with no semantic value of its own, must be referentially linked to an antecedent that does. In both language and music, FI thus performs a "filtering function" of the sort discussed in section 3.1. Because of FI, a structure that is formally derivable by Merge may nonetheless be perceived as deviant because it contains one or more elements to which the interpretive systems cannot assign an interpretation.

This is one of the more important respects in which the architecture of our model differs from the architecture of the GTTM model, which contains no component with a filtering function.[57] GTTM does treat cadenced structures as special — but only as a parsing preference. One of the preference

---

[56] A similar idea was suggested in passing by Steedman (1984, 56) — a point elaborated on in a later paper by the same author (1994, 125), which suggests that the "structural rules" of musical syntax are related by "rules of interpretation" to a "semantics" one of whose tasks is to "compositionally define key in terms of cadences, or progressions of chord tonalities".

[57] A similar point is made by Sundberg & Lindblom (1991, 260), in their overview of a variety of proposed generative theories of musical genres — grammars of the kind referred to here as type-2. They note that "…[Lerdahl & Jackendoff] do not address the issue of describing musical style in the sense of what is possible and what is impossible." They go on to suggest (pp. 261-262) that such a description is both worthwhile and feasible, and that GTTM's insistence on the futility of such a task is overly pessimistic.

rules of their system (*Prolongational Reduction Preference Rule 6*, p. 234) dictates that a listener faced with an ambiguous musical surface that could be analyzed either with or without a final cadence will prefer the analysis that includes the cadence. Though GTTM calls the favored PR configuration "Normative Prolongational Structure", the notion of "normativity" is actually alien to the logic of the GTTM model. Similarly, despite the fact that the notion of Normative Prolongational Structure is introduced as a response to the fact that "certain tree patterns recur constantly, whereas others virtually never happen" (GTTM 197), the GTTM model is actually incapable of explaining this fact — precisely because it lacks any way of enforcing the presence or absence of a particular structural configuration, in contrast to the model sketched here.

In response, one might object to our proposal that a cadence is crucial to interpretation on empirical grounds. After all, it is often possible for a listener to tell what key a piece is in even without hearing a full cadence in that key — for example, if a piece terminating in a full cadence is interrupted before that cadence. In general, since no two keys (except for relative major-minor pairs) share more than four chords, any five distinct chords chosen from a single diatonic collection (i.e. from the pitches allowed in a given key) suffice to uniquely identify the collection from which they are chosen, and for most key comparisons fewer than five chords will suffice. Nonetheless, in tonal music up to the 19th century, it is the cadence — not merely the sounding of a sufficient number of tones — that establishes the key in a manner felt to be satisfying, proper, or well-formed.[58] On this view, the observation that a listener can nevertheless infer the current key without hearing its cadence has a status similar to the observation that we can often assign an interpretation to a linguistic sentence that is ill-formed or incomplete. Consider the utterances in (63):

(63) a. "the boy said that Sue criticized himself"
b. "John goes to school every"

Utterance (63a) violates principle A of the Binding Theory, and is felt to be deviant. Nonetheless, despite its deviance, English speakers have no trouble narrowing down the intended meaning of the utterance to one of two possibilities. Similarly, although utterance (63b) is missing one or more words at the end, speakers assign it a meaning and strong expectations about how it should end. Although such utterances are interpretable to various degrees, and do occur (when speakers make errors or noise interferes with perception of a speech signal), we never lose the clear intuition that they are deviant when evaluated as sentences of English. Consequently, the right theory of grammar should characterize them as such. We believe that the status of a piece of tonal music without a cadence is similar. Although it is often possible to infer key from a variety of clues, a listener who does not hear the cadence has an intuition that something is "incomplete" — the key has not been properly established.

If the theory developed in the preceding sections is correct, the phenomenon of *cadence* in music exemplifies a number of specific and architectural similarities between music and language. The specific properties of cadences suggest an operation of head-movement similar to that found robustly in the syntax of languages. The role of the cadence in establishing a key, and the formal details of this

---

58 As Jean-Philippe Rameau (1722) noted in the second volume of his *Traité de l'harmonie*: "L'On appelle Cadence parfaite, une certaine conclusion de chant, qui satisfait de façon, que l'on n'a plus rien à désirer après une telle Cadence." ["The perfect cadence is a certain way of ending a strain which is so satisfying that we desire nothing further after it.", transl. Gossett (1971, 63)] No such intuition arises when one has merely deduced the current key from a collection of chords or pitches without a cadence.

role, motivated a distinction between the syntax of music proper (PR) and a Tonal-Harmonic interpretive component (THC) that mirrors the interaction of syntax and semantics within language. Furthermore, we have seen some reason to suspect that music, like language, contains a principle of Full Interpretation, which requires each object produced within PR to be properly interpreted by THC, much as each syntactic object in language must be interpretable by the semantic component.

6.2 Comparison with Rohrmeier's (2008) model

In the previous section, we argued that key membership and harmonic properties are determined in an interpretive THC component separate from the narrow syntax. This section elaborates on the arguments for this analytical move, and compares this system to one with harmonic function encoded in syntactic structure.

Since domination is irreflexive, a Key Domain itself does not necessarily belong to the key that it establishes. For example, a tonic-marked G-major constituent α is not necessarily in G-major, but is in whatever key is established by a higher Key Domain that dominates α. This property of the system was demonstrated in (58b) where the tonic-marked G-major chord 3 served as the tonic of the key of G-major at the node G′, while its maximal projection (GP) served as a dominant of C-major. This fact reinforces the validity of a fundamental property of the PR structures posited in GTTM, which we have preserved in our presentation as well: *what projects syntactically is pitch-class*. Crucially, scale-degrees (e.g. those traditionally represented by Roman-numerals such as V or I) and harmonic functions (dominant, tonic, etc.) do not project in this fashion. If they did, we would have to ask why a dominant chord may also function as a tonic, and why a V-chord may simultaneously function as a I. By contrast, a G is always a G, at every relevant level of PR structure.

This means, however, that if our goal is a type-2 theory of music, the properties of music that we have just discussed — the requirement of a cadence as a key-establishing device, and the requirement of a key-establishment for interpretation in THC — cannot be built directly into the syntax of PR. If the syntax does not contain categories like *tonic* and *V*, requiring these categories to stand in a particular configuration can not be accomplished in the syntax. In this context, it is instructive to compare our approach to another proposal whose overall goals are similar to ours — but which differs on precisely this point.

Rohrmeier (2008) develops a grammar fragment for harmonic structure whose representations, like those of GTTM and the present paper, are tree-structures (acyclic directed graphs) reminiscent of proposals for language syntax. These structures are generated by a phrase-structure grammar like those familiar from early generative syntax. Crucially, however, information about scale-degree and harmonic function are built into syntactic structure directly, and the obligatoriness of a full or half cadence (in any piece containing more than just a sustained or repeated tonic) is similarly built into the 'rewrite rules' that expand the initial symbol *t* ('tonic'):[59]

---

59 In these respects, Rohrmeier's proposal is broadly similar to the 12-bar blues grammar proposed by Steedman (1984).

(64) **Phrase-structure rewrite rules from Rohrmeier (2008) (simplified, omitting special rules for minor keys and other details)**

a.	t → t t	e.	t → I
b.	t → t d	f.	d → {V, VII}
c.	t → d t	g.	s → {IV, II, VI}
d.	d → s d		

This system captures harmonic generalizations, such as the obligatoriness of a cadence, by including several layers of information in syntactic trees. For instance, rules (64a-c) entail that a piece consists minimally of a series of tonics; that the next-most-complicated piece will intersperse dominants between these tonics; and that all more complicated pieces will also include this basic structure. This proposal ensures that all but the most trivial pieces will conclude with either a dominant-tonic sequence (a full cadence) or a tonic-dominant (half-cadence) sequence (which we have not discussed here), possibly followed by a prolongation of the final cadential element. Rules (64e -f) further ensure that chord progressions corresponding to these final sequences will be comprised of I, V, and vii.

These rules contain several levels of information because fully-interpreted pitch events themselves involve several levels of information. For instance, a G chord in the key of C is composed of the pitches G-B-D; belongs to the key of C; is the V chord (built on the fifth scale degree) in this key; and is a chord with a dominant function. At the same time, these four types of information are not independent: any two may be derived from the other two. Another way of saying this is that some levels of harmonic information are predictable from other levels; an adequate theory of harmony must characterize the *relationships between levels of harmonic information*.

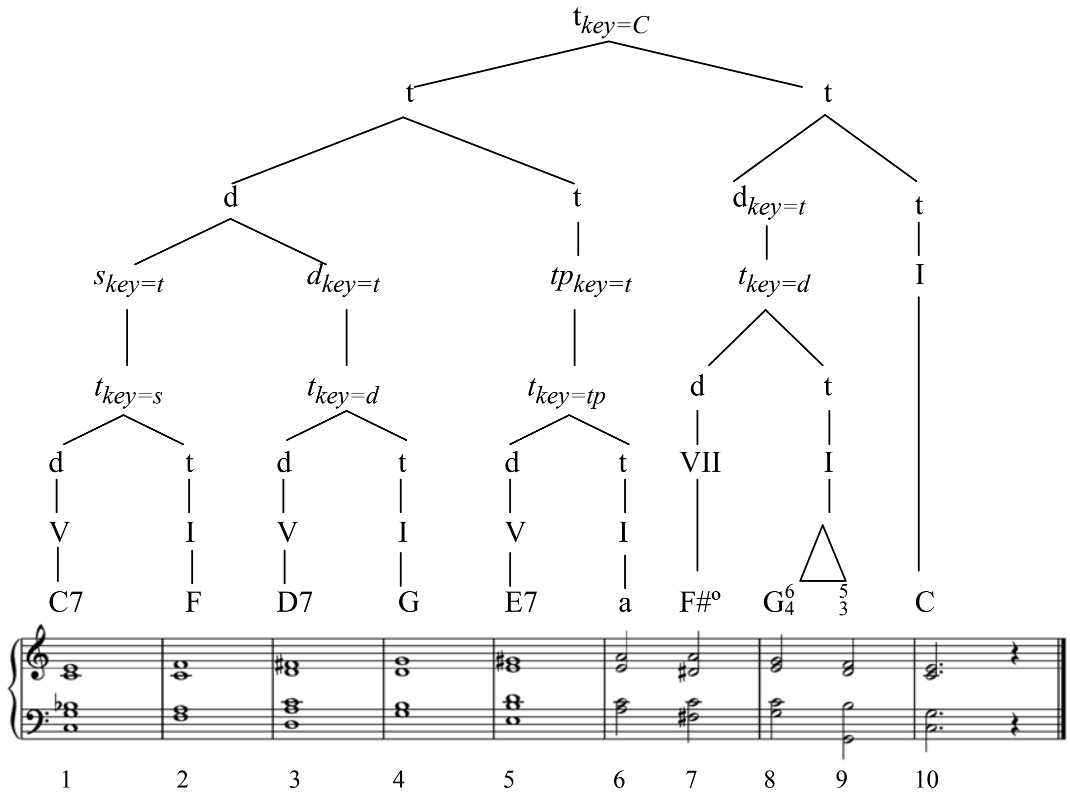
Rohrmeier tackles this problem in two ways:

1. the relationship of pitch class to key and scale-degree (e.g. "V") is indicated by diacritics on node labels (a notation that suggests that the work of figuring out how pitches relate to key areas is done elsewhere in the grammar); but

2. the relationship of scale-degree to function is handled directly in the phrase structure grammar, as in (64e-g) above.

By listing all allowable chord progressions in terms of their functional labels, and then listing all of the chords that may fulfill each function, we come to a generative characterization of the relationship between chords, their functions, and the contexts they may appear in. The types of structure generated by this proposal can be see in a passage from Bortnyansky's choral piece *Tebe poëm*, for which Rohrmeier proposes the harmonic structure in (65):

(65) **Structure assigned by Rohrmeier to *Tebe poëm* excerpt**[60]



The choice of *t* as the head of the various [*d t*] constituents in (65) captures the fact that the passage reduces to the progression F - G - a - G - C (i.e. IV - V - vi - V - I). Each chord in this progression is immediately preceded in the musical surface by a dominant of its own, attached as a dependent left branch. In this system, unlike ours, "Tonic" and "Dominant" (*t* and *d*) are syntactic categories that determine the harmonic functions of chords that they dominate.[61] A chord that bears more than one harmonic function, like the first G in (65) (chord 4), must therefore do so by virtue of being dominated by more than one harmonic-function node. This G is characterized as both a local tonic and a higher dominant because it is dominated by the category *t* at a low level of structure and by the category *d* at a higher level.

The idea that harmonic functions are syntactic categories contrasts with the GTTM-based approach developed here. In our approach, as in GTTM, the elements that project higher levels of harmonic structure are pitch classes themselves, not harmonic functions, and harmonic function is not

---

60 In full:

61 "Harmonic function" in Rohrmeier's proposal amounts to the classification of chords imposed by sets of phrase structure rules. By virtue of rule (64f), for example, V and vii may both occur as terminals wherever other rules license the occurrence of the non-terminal node "d". This is how the notion "V and vii may instantiate the dominant function" is reconstructed in this approach.

represented in PR, but is a by-product of the distinct system THC. We believe that the properties of examples like (65) actually argue in favor of our approach, and against the purely syntactic view of harmonic function advanced by Rohrmeier. We also find the comparison instructive, since it highlights an advantage of GTTM's pitch-based PR that has not so far been made explicit (either in the present paper or in GTTM itself).

Consider how Rohrmeier's grammar fragment enforces the presence of the G we have been discussing (chord 4) as the terminal node dominated by the head of $t_{key=d}$. The grammar accomplishes this through a chain of rules whose structural result is a chain of immediate dominance relations: *t* may immediately dominate [*d t*] by rule (64c); *t* may dominate scale-degree I by rule (64e). Though Rohrmeier is not specific on this point, some other component of the grammar that computes pitch spaces will presumably determine that the I-chord in G-major dominates a chord comprised of pitch classes G, B, and D. The fact that chord 4 fulfills the tonic function in G-major, even though it is not immediately dominated by the G-major node, is captured straightforwardly by a chain of immediate domination relations that first identifies scale-degree I as the instantiation of tonic and then identifies the G-major chord as the instantiation of I in G-major.

Rohrmeier's grammar does not explicitly allow for the immediate domination of $t_{key=d}$ (the "G major node") by $d_{key=t}$ (a "C major node") nor for the various other "key switching" relations of immediate dominance characteristic of the continuously modulating (65). Whatever rules might license such relations in a more complete specification of the grammar, it will be crucial to make sure that the same chord (the same pitch classes) that successfully constitute the tonic in G are also appropriate as a dominant in C. How can this be done? All things being equal, we expect that the way in which the system decides which chord may function as the dominant in C in *Tebe poëm* should be the same as the way in which the system decides which chord may constitute a tonic, dominant or subdominant in G or any other key. In fact, however, this is not possible in Rohrmeier's system. Nothing like the chain of immediate dominance relations that accomplishes this task for the "G major node" in (65) can be invoked to accomplish the parallel task for the "C major node" — precisely because the C-major node immediately dominates the G-major node rather than a node like *d* which would in turn immediately dominate a scale-degree such as V (in turn licensing the terminal node G). Some additional non-local special mechanism must be postulated instead, to allow the C-major node to verify that the terminal node G may in fact function as its dominant, despite the lack of a chain of immediate dominance connecting the C-major node to *t*, *t* to I, and I to G.[62] As a consequence, what should clearly be a uniform procedure for deciding the possible harmonic functions of a given chord will need to be handled in two rather different ways.

Crucially, this problem arises in Rohrmeier's grammar precisely because headedness in this system is a property of harmonic function rather than pitch-class.[63] If pitch-class information propagates up the tree as in GTTM, the information that a given chord α is an instance of G is available

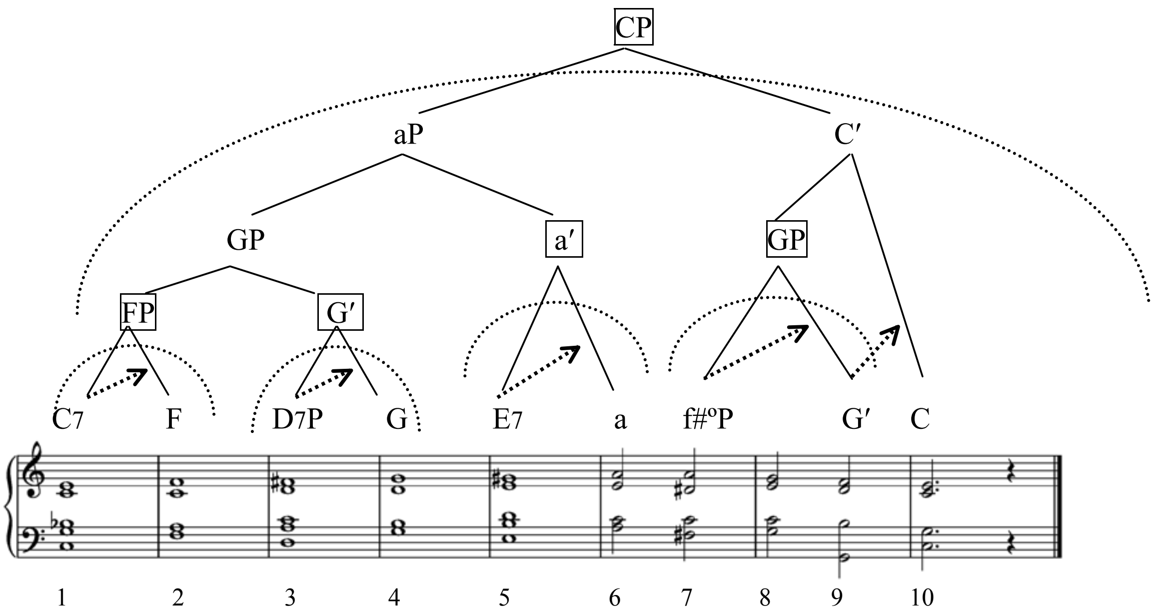
---

[62] Alternatively, a separate computation might determine that $t_{key=d}$ may immediately dominate $d_{key=t}$ on the grounds that when the properties of tonal pitch space are consulted, the tonic of the dominant and the dominant of the tonic may be instantiated by the same chord. In Rohrmeier's system however, even this move would fail unless supplemented with the requirement that they actually *are* instantiated by the same chord in the relevant structure, since the dominant of the tonic has more than one possible instantiation (V and vii), only one of which may also function as the tonic of the dominant.

[63] As far as we can tell, the (HPSG-inspired) grammar fragment for chord progression and cadence in Tojo et al. (2006) has the same property, raising the same questions.

at every level of structure at which α remains the head. Consequently, the harmonic function of any chord can always be determined locally even in circumstances of key-switching, so long as the current key is known to the system. This is shown in (66), where the dotted arrows indicate tonic-marking head-movement, the boxed nodes indicate key-domains and the dotted umbrella lines serve as a reminder that it is nodes dominated by a key domain α that count as "in the key of α", as required by (56) and (59):

(66) **PR for *Tebe poëm* excerpt in framework of this paper**



Recall finally that we have already argued against the one obvious alternative to this conclusion, namely that *both* pitch-class and harmonic function information propagate syntactically. As we saw in the previous section, this alternative fails precisely because harmonic function information does not extend past key boundaries, whereas pitch-class information does. If both propagate syntactically, we would need two entirely separate phrase-structure grammars to characterize this difference.

In this section, we have compared our proposal, in which the notions *key* and *harmonic function* are properties of a separate interpretive component, to an alternative proposal in which they are built into the syntax. The behavior of chords that function in more than one key favors a system of the sort we have proposed over the alternative.

7 <u>Conclusions</u>

In this paper we proposed and explored some consequences of the Identity Thesis in (1):

(1) **Identity Thesis for Language and Music**
All formal differences between language and music are a consequence of differences in their fundamental building blocks (arbitrary pairings of sound and meaning in the case of language; pitch-classes and pitch-class combinations in the case of music). In all other respects, language and music are identical.

We then identified certain specific properties of musical and linguistic structure that (when properly aligned) appear to coincide in their formal details — thus supporting the thesis.[64] In section 4, we argued that linguistic syntax and prosody are formally identical to musical PR and TSR, respectively — and that the *interface* between these two components is also formally similar (perhaps identical) in music and language. In section 5, we argued that the cluster of phenomena known as *head-movement* in language is formally identical to the cluster of phenomena known as *cadence* in music. Finally, in section 6, we laid out the arguments for positing a separate interpretive component for musical structures, THC: the counterpart for music of the semantic component for language.

More generally, our investigations suggest certain broader methodological points about how inquiry into musical and linguistic structures might fruitfully proceed. The first of these concerns the level of detail that one must achieve in analyzing both music and language before one can pursue useful comparisons between the two domains.

Given the uncontroversial fact that language has a lexicon but music does not, research must deal with the two domains at a considerable level of abstraction — in order to uncover whatever parallels exist. This in turn means that research must simultaneously provide both detailed investigations of the data that motivate abstract analyses and detailed investigations of the formal properties of the analyses themselves. After all, in order to have any hope of discovering similarities between the formal systems that model linguistic and musical structures, the systems themselves must be correct, and must be fully specified.

---

64 The view of linguistic syntax that we have presupposed in this paper reflects the research tradition often referred to as "Principles and Parameters" (which in turn incorporates specific ideas developed in the 1980s and early 1990s under the name "Government-Binding Theory" — and more recently, as the "Minimalist Program"). A number of other approaches to syntax are also under active investigation, which differ from Principles and Parameters syntax to varying degrees in goals, research philosophy, and technical details. To the extent that this paper has argued for the relevance to music of specific points of linguistic analysis that distinguish one approach from another, this paper itself can be said to make a contribution to corresponding debates within linguistics. Research in *Head-Driven Phrase Structure Grammar* (HPSG), for example, has developed notions reminiscent of External Merge, generating headed (though not necessarily binary-branching) structures for language not too dissimilar from those discussed here (Pollard & Sag 1994; Sag, Wasow & Bender 2003). Internal Merge, however, does not form part of the repertoire of possibilities generally considered in HPSG literature. In particular, though recognizable analogues to Internal Merge have been developed to explain such phenomena as *wh*-movement, a very different treatment has generally been suggested by HPSG researchers for phenomena like those that we have treated as cases of head movement within language (Pollard & Sag 1994, 40-43; Kim & Sag 2002).

We will not attempt here to evaluate the linguistic issues at stake, but merely note that if the Identity Thesis is correct, evidence for or against head movement in *music* could in principle play a decisive role in adjudicating among competing proposals concerning language. If it should turn out, for example, that the HPSG mechanisms proposed as an account of phenomena otherwise treated as head movement have no role to play in music (or conversely, that the mechanisms necessary to explain the δ-τ adjacency requirement in musical cadences have no counterpart in HPSG's linguistic proposals), the result might be a novel argument favoring head movement over its HPSG counterpart for language. An opposite result might support opposite conclusions, of course — or might reveal hitherto undetected inadequacies in both sets of proposals.

These points are familiar from the study of linguistic typology. There, a distinction is often made (Chomsky 1965, 28ff.) between *substantive universals* (items that all languages share or lack) and *formal universals* (abstract structural properties that all languages share or lack). Though substantive universals are few, formal universals are richly attested and constitute the foundation of most successful current research on the structure of language.[65]

This lesson applies to the search for music-language parallels as well. In order to uncover such parallels, it is not enough to scan the surface of musical and linguistic objects and pick out items that resemble each other. Nothing in the musical or linguistic surface, for instance, would suggest that the cognitive representation of tonal and rhythmic prominence in music shares the same formal mechanism as that of phrasing and stress in language. Uncovering that parallel required many years of cross-linguistic research on prosody, on the one hand, and the theoretical advances made by GTTM, on the other. The comparison would not have been possible without the labor of many linguists and many music theorists whose efforts have yielded detailed, explicit theories of musical and linguistic structure.

Similarly, nothing obvious about verbal morphosyntax cries out for a comparison with cadences. Rather, certain formal characteristics of cadences stand out as exceptional in the context of an explicit theory such as GTTM. Cadences feature an adjacency requirement; they result in a tight coupling of two elements; the two elements still behave as if they are structurally independent; and the phenomenon is required in order to build a syntactic structure that can be interpreted. These same formal characteristics are familiar to linguists from over 20 years of research on head-movement. As a result, a deep parallel emerges between music and language, a parallel that would have remained obscure in the absence of explicit, pre-existing formal theories in the two domains.

A further methodological point was elaborated in section 2. In order to compare two explicit theories, we must first align them. Because GTTM is cast as a *generative parser*, in our terms, and most linguistic theories are *generative grammars*, similarities between the two theories may be hidden.

The system of mapping from TSR to PR in music — cast in GTTM as a derivational algorithm that works its way down from the root of a tree to its terminal nodes — has little in common with pre-existing characterizations of the mapping from syntax to prosody in language. Nonetheless, it proved possible to recast theories of both linguistic and musical interfaces in a 'direction-neutral' way, by positing constraints on the overall form of the relationship between two components of grammar. When this was done for the PR-TSR interface and the syntax-prosody interface, it turned out that the only significant difference was the specific content of the Region Condition applicable to each domain. (We suggested that even this difference may be illusory.) Both interfaces translate between a representation of syntactic relations and a representation of prominence; both require some degree of isomorphy between the two types of structure; and both allow limited departures from this isomorphy. The apparent differences between the two systems can be traced, again, to the differences between the basic building blocks of language and those of music. Discovering this parallel required recasting both theories so that they were commensurate in goals and approach. Comparing a type-I theory to a type-II one would not have allowed these similarities to emerge.

---

65 See Baker (2001) for a non-technical survey and Newmeyer (2005) for more detailed discussion. Most generative research on linguistic structure concerns itself with the discovery of formal universals or the exploration of their consequences for the grammar of individual languages, whether implicitly (as when findings from one language are brought to bear in research on another) or explicitly (as when explicit cross-linguistic comparison is carried out. See, for example the papers collected in Cinque & Kayne (2005) for recent examples of both strategies.

A final methodological point concerns the Identity Thesis itself. We hope to have demonstrated the fruitfulness of this hypothesis as a starting point for investigation — precisely because it is the strongest plausible hypothesis about the relationship between language and music. Only when every attempt to understand some set of musical and linguistic data in conformity with the Identity Thesis has been exhausted can we be sure that the differences that we have uncovered are real. We may then proceed to ask *why* such differences might exist.

An alternative strategy might start with the differences — and ask as its *first* question why these differences exist. In our view, however, this question is unanswerable in the absence of a theory of the cognitive faculties *shared* by the two domains. The observation that language has nouns and verbs, while music has notes and chords, is not particularly edifying in and of itself. It becomes an important observation when it is combined with an explicit hypothesis about what is *not* different between language and music. As we attempted to show in this paper, the basic difference in building blocks may then be used to derive and explain a number of far-ranging and profound differences between language and music.

If correct, the Identity Thesis raises a profound question for which we currently have no answer whatsoever, namely: *why* is it correct? Why do such different cognitive entities as lexical items and pitch-classes combine and recombine by Merge (both External and Internal), forming structures that interact in identical (or nearly identical) ways with adjacent components of grammar?

This question appears at first glance to bear on ancient debates about "domain-specificity". If one presupposes that language and music are distinct cognitive faculties, then one might take evidence that bears on the Identity Thesis to also bear on the "domain-specificity" of Merge and the grammatical components that interact with it. We know of no particular reason, however, to accept the presupposition that language and music constitute distinct faculties of mind *en bloc*. If our Identity Thesis is correct, music and language are the same mental faculty in one sense (formal resources), and distinct mental faculties in another (building blocks) — and whether this conclusion should be understood as a blow for or against "domain-specificity" or not is in the end a terminological matter of no scientific interest . Likewise, it does not follow from the proposition that music and language share particular formal resources that these resources are also relevant *outside* music and language. This, like the issues taken up in this paper, is ultimately an empirical matter — and once again, it could turn out that the answer is not unequivocal or simple. See Giblin (2008) for extensive insightful discussion of these and related issues.

Finally, we hope that this paper serves as a demonstration that comparison between musical and linguistic structure can proceed in a rigorous and principled fashion. As we noted at the outset, serious structural comparison of the sort begun here has been undertaken far less often than one might have hoped — especially when compared to the flourishing research programs that contrast music and language from the perspectives of neuroscience and cognitive psychology. This may be partially due to the degree of specialization that research in either linguistics or music theory demands. Work in either field requires specialized knowledge, mastery of particular formalisms and notational conventions, and a substantial technical vocabulary. To our minds, however, these factors simply reflect the fact that both fields have independently developed a large body of remarkable results — which is what makes comparative work across the two domains so rewarding and potentially important to both fields. The Identity Thesis in particular suggests that informed dialogue between music theorists and linguists can help *both* fields to advance.

REFERENCES

Baker, Mark C. 2001. *The atoms of language*. New York: Basic Books.

Balzano, Gerald J. 1982. The pitch set as a level of description for studying musical pitch perception. In Manfred Clines, ed. *Music, mind and brain, the neurophysiology of music*. 321-351. New York: Plenum Press.

Bernstein, Leonard. 1976. *The Unanswered Question : Six talks at Harvard*. Cambridge, Mass: Harvard University Press.

Bigand, Emmanuel. 1990. Abstraction of two forms of underlying structure in a tonal melody. *Psychology of Music.* 18:45. [http://pom.sagepub.com/cgi/content/abstract/18/1/45]

Bigand, Emmanuel, and Bénédicte Poulin-Charronnat. 2006. Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition.* 100:100-130. [http://linkinghub.elsevier.com/retrieve/pii/S0010027705002234]

Bresnan, Joan W. 1971. Sentence stress and syntactic transformations. *Language.* 47:257-281. [http://www.jstor.org/stable/412081]

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1986a. *Barriers*. Cambridge, Mass: MIT Press.

Chomsky, Noam. 1986b. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Chomsky, Noam. 1995a. Bare phrase structure. In Gert Webelhuth, ed. *Government binding theory and the minimalist program*. 383-439. Oxford: Oxford University Press.

Chomsky, Noam. 1995b. *The minimalist program*. Cambridge: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries. In Roger Martin, David Michaels, and Juan Uriagereka, eds. *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*. 89-156. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz, ed. *Ken Hale : A life in language*. 1-52. Cambridge, Mass.: MIT Press. [http://cognet.mit.edu/library/books/view?isbn=0262112574]

Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Balletic, ed. *Structures and beyond: The cartography of syntactic structures, vol. 3*. 104-131. Oxford: Oxford University Press. [first published as *MIT occasional papers in linguistics: vol. 20 (*2001), Cambridge, MA :MIT]

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry.* 36:1–22.

Chomsky, Noam. 1980. On binding. *Linguistic Inquiry.* 11:1-46. [http://www.jstor.org/stable/4178149]

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row, Publishers.

Cinque, Guglielmo, and Richard S Kayne. 2005. *The Oxford handbook of comparative syntax*. Oxford ; New York: Oxford University Press.

Clifton, Charles, and Adrian Staub. 2008. Parallelism and competition in syntactic ambiguity resolution. *Language and Linguistics Compass.* 2:234-250. [http://ciillibrary.org:8000/ciil/Fulltext/Language_and_Linguistic_Compass/2008/Vol_2_2_2008/Article_1.pdf]

Emonds, Joseph. 1978. The verbal complex v′ - V in French. *Linguistic Inquiry.* 9:151-175. [http://www.jstor.org/stable/4178050]

Fedorenko, Evelina, Aniruddh Patel, Daniel Casasanto, Jonathan Winawer, and Edward Gibson. 2009. Structural integration in language and music: Evidence for a shared system. *Memory and Cognition.* 37:1-9. [http://mc.psychonomic-journals.org/content/37/1/1.full.pdf+html]

Fitch, W. Tecumseh. 2006. The biology and evolution of music: A comparative perspective. *Cognition.* 100:173-215. [http://linkinghub.elsevier.com/retrieve/pii/S0010027705002258]

Forte, Allen. 1959. Schenker's conception of musical structure. *Journal of Music Theory.* 3:1-30. [http://www.jstor.org/stable/842996]

Fox, Danny. 1995. Economy and scope. *Natural Language Semantics.* 3:283-341.

Fox, Danny. 1999. *Economy and semantic interpretation*. Cambridge, Mass.: MIT Press.

Fukui, Naoki, and Hiromu Sakai. 2003. The visibility guideline for functional categories: Verb raising in Japanese and related issues. *Lingua.* 113:321-375. [http://linkinghub.elsevier.com/retrieve/pii/S0024384102000803]

Giblin, Iain. 2008. *Music and the generative enterprise: Situating a generative theory of tonal music in the cognitive sciences*. Doctoral dissertation, University of New South Wales.

Ginzburg, Jonathan, and Ivan A Sag. 2002. *English interrogative constructions*. Stanford: CSLI.

Halle, John, and Fred Lerdahl. 1993. A generative textsetting model. *Current Musicology.* 55:3-23.

Hauser, Marc D., and Josh McDermott. 2003. The evolution of the music faculty: A comparative perspective. *Nature Neuroscience.* 6:663-668.

Hayes, Bruce. 1989. The prosodic hierarchy in meter. In Paul Kiparsky, and Gilbert Youmans, eds. *Phonetics and phonology, volume 1: Rhythm and meter*. 201-260. Orlando, FL: Academic Press.

Hayes, Bruce. 1995. *Metrical stress theory : Principles and case studies*. Chicago: University of Chicago Press.

Hayes, Bruce. 2009. Textsetting as constraint conflict. In Jean-Louis Aroui, and Andy Arleo, eds. *Towards a typology of poetic forms: From language to metrics and beyond*. Amsterdam: Benjamins. [http://www.linguistics.ucla.edu/people/hayes/textsetting/]

Hayes, Bruce, and Margaret R. MacEachern. 1996. Are there lines in folk poetry? *UCLA Working Papers in Phonology.* 1:125-142.

Huang, C. T. James. 1981. Move wh in a language without wh movement. *The Linguistic Review.* 1:369-416.

Hughes, David W. 1991. Grammars of non-western musics: A selective survey. In Peter Howell, Robert West, and Ian Cross, eds. *Representing musical structure*. 327-362. London and San Diego: Academic Press.

Huron, David. 2003. Is music an evolutionary adaptation? In Isabelle Peretz, and Robert J Zatorre, eds. *3(1)*. 57-75. Oxford ; New York: Oxford University Press.

Jackendoff, Ray. 1977. Review of The Unanswered Question by Leonard Bernstein. *Language.* 53(4):883-894. [http://www.jstor.org/stable/412916]

Jackendoff, Ray. 1987. *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 1992. *Languages of the mind : Essays on mental representation*. Cambridge, Mass.: MIT Press.

Jackendoff, Ray. 2009. Parallels and nonparallels between language and music. *Music Perception.* 26:195-204.

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

Kaisse, Ellen M. 1985. *Connected speech : The interaction of syntax and phonology*. Orlando: Academic Press.

Katz, Jonah. 2008. Towards a generative theory of hip-hop. Talk delivered at *Music, Language and the Mind.* Tufts University. [http://web.mit.edu/jikatz/www/KatzGTHH.pdf]

Katz, Jonah. 2007. Optimal musical grouping. Talk delivered at *Language and Music as Cognitive Systems.* Cambridge, UK.

Kayne, Richard S. 1984. Unambiguous paths. In Richard S. Kayne, ed. *Connectedness and binary branching*. 129-164. Dordrecht: Foris Publications.

Kim, Jong-Bok, and Ivan A. Sag. 2002. Negation without head-movement. *Natural Language & Linguistic Theory.* 20:339-412. [http://www.jstor.org/stable/4048030; http://ftp-linguistics.stanford.edu/sag/nllt-ks/kim-sag-nllt.pdf]

Kiparsky, Paul. 2006. A modular metric for folk verse. In B. Elan Dresher, and Nila Friedberg, eds. *Formal approaches to poetry: Recent developments in metrics*. Berlin: Mouton de Gruyter. [http://www.stanford.edu/~kiparsky/Papers/hayes.pdf]

Kishimoto, Hideki. 2007. Negative scope and head raising in Japanese. *Lingua.* 117:247-288. [http://linkinghub.elsevier.com/retrieve/pii/S0024384106000283]

Kishimoto, Hideki. 2008. On verb raising. In Shigeru Miyagawa, and Mamoru Saito, eds. *The Oxford handbook of Japanese linguistics*. 107-140. New York: Oxford University Press. [http://www.loc.gov/catdir/toc/ecip082/2007041899.html]

Kisseberth, Charles W., and Mohammad Imam Abasheikh. 1974. Vowel length in Chi-mwi:ni: A case study of the role of grammar in phonology. In Anthony Bruck, Robert Allen Fox, and Galy La, Michael W., eds. 193-201. Chicago Linguistic Society.

Koizumi, Masatoshi. 1995. *Phrase structure in minimalist syntax*. Doctoral dissertation, Massachusetts Institute of Technology.

Koizumi, Masatoshi. 2000. String vacuous overt verb raising. *Journal of East Asian Linguistics.* 9:227-285. [http://www.springerlink.com/index/G867942WW8HG366K.pdf]

Koopman, Hilda J., and Anna Szabolcsi. 2000. *Verbal complexes*. Cambridge, Mass.: MIT Press.

Krumhansl, Carol L. 1990. *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Larson, Richard. 1988. On the double object construction. *Linguistic Inquiry.* 19:335-392.

Lerdahl, Fred. 2001a. The sounds of poetry viewed as music. *Annals of the New York Academy of Sciences.* 930:337-354. [http://www3.interscience.wiley.com/journal/118979182/abstract]

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

Lerdahl, Fred. 2004. A music-theoretic approach to the sounds of poetry. Unpublished manuscript. Columbia University.

Lerdahl, Fred. 2009. Genesis and architecture of the GTTM project. *Music Perception.* 28:187-194.

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

Lerdahl, Fred, and Carol Krumhansl. 2007. Modeling tonal tension. *Music Perception.* 24:329-366.

Liberman, Mark, and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry.* 8:249-336. [http://www.jstor.org/stable/4177987]

Marvin, Tatjana. 2003. *Topics in the stress and syntax of words*. Doctoral dissertation, Massachusetts Institute of Technology. [http://dspace.mit.edu/handle/1721.1/7988]

Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry.* 37:69–109.

Michaels, Jennifer. 2009. To alternate or not to alternate: What is the boundary? In *Proceedings of NELS 40*. GLSA. [http://web.mit.edu/jenmich/www/NELS_Michaels.pdf]

Nespor, Marina, and Irene Vogel. 1986. *Prosodic phonology*. Dordrecht, Holland ; Riverton, N.J.: Foris.

Newmeyer, Frederick J. 2005. *Possible and probable languages : A generative perspective on linguistic typology*. Oxford ; New York: Oxford University Press.

Otani, Kazuyo, and John Whitman. 1991. V-raising and VP-ellipsis. *Linguistic Inquiry.* 22:345-358. [http://www.jstor.org/stable/4178724]

Patel, Aniruddh D. 2003. Language, music, syntax and the brain. *Nature Neuroscience.* 6:674-681. [http://www.nature.com/neuro/journal/v6/n7/abs/nn1082.html]

Patel, Aniruddh D. 2008. *Music, language, and the brain*. Oxford ; New York: Oxford University Press. [http://www.loc.gov/catdir/toc/ecip0715/2007014189.html]

Patel, Aniruddh D. 2006. Musical rhythm, linguistic rhythm, and human evolution. *Music Perception.* 24:99-104.

Peretz, Isabelle. 2006. The nature of music from a biological perspective. *Cognition.* 100:1-32. [http://dx.doi.org/10.1016/j.cognition.2005.11.004]

Peretz, Isabelle. to appear. Music, language and modularity in action. In Patrick Rebuschat, Martin Rohrmeier, John Hawkins, and Ian Cross, eds. *Language and music as cognitive systems*. 57-75. Oxford ; New York: Oxford University Press. [http://www.isc.uqam.ca/docs/Peretz_LanguageAndMusic2008.pdf]

Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In Eric J. Reuland, and Alice G. B. ter Meulen, eds. *The representation of (in)definiteness*. Cambridge, Mass.: MIT Press.

Pesetsky, David. 2000. *Phrasal movement and its kin*. Cambridge, Mass.: MIT Press.

Pesetsky, David. in prep. Same recipe, different ingredients: Music syntax is language syntax. Unpublished manuscript.

Pesetsky, David, and Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In S. Karimi, V. Samiian, and W. Wilkins, eds. *Phrasal and clausal architecture*. Amsterdam: Benjamins. [http://web.mit.edu/linguistics/www/pesetsky/Pesetsky_Torrego_Agree_paper.pdf [written 2004]]

Pinker, Steven. 1997. *How the mind works*. New York: Norton.

Pollard, Carl Jesse, and Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago: CSLI/University of Chicago Press.

Pollock, Jean-Yves. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry.* 20:365-424. [http://www.jstor.org/stable/4178634]

Prince, Alan, and Paul Smolensky. 2004. *Optimality theory : Constraint interaction in generative grammar*. Malden, MA: Blackwell Pub. [http://www.loc.gov/catdir/toc/ecip0413/2004000723.html]

Prince, Alan S., and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Unpublished manuscript. Rutgers University Center for Cognitive Science. Piscataway, New Jersey.

Rameau, Jean-Philippe. 1722. *Traité de l'harmonie réduite à ses principes naturels*. Paris: Ballard. [http://www.chmtl.indiana.edu/tfm/18th/RAMTRA2_TEXT.html]

Rameau, Jean-Philippe. 1971. *Treatise on harmony*. New York: Dover Publications.

Richards, Norvin. 1997. *What moves where when in which languages?*

Richards, Norvin. 2001. *Movement in language: Interactions and architectures*. Oxford: Oxford University Press.

Rohrmeier, Martin. 2008. A generative grammar approach to diatonic harmonic structure. In Haralambos Spyridis, Anastasia Georgaki, Georgios Kouroupetroglou, and Christina Anagnostopoulou, eds. *Proceedings of the 4th sound and music computing conference*. 97-100. [http://linkinghub.elsevier.com/retrieve/pii/S0010027705000867]

Rouveret, Alain, and Jean-Roger Vergnaud. 1980. Specifying reference to the subject: French causatives and conditions on representations. *Linguistic Inquiry.* 11:97-202. [http://www.jstor.org/stable/4178152]

Sag, Ivan A., Thomas Wasow, and Emily M. Bender. 2003. *Syntactic theory: A formal introduction, second edition*. Chicago: CSLI Publications [University of Chicago Press].

Schenker, Heinrich, and Felix Salzer. 1932. *Five graphic music analyses (Fünf Urlinie-Tafeln)*. New York: Dover Publications.

Selkirk, Elisabeth O. 1972. *The phrase phonology of English and French*. Doctoral dissertation, Massachusetts Institute of Technology.

Selkirk, Elisabeth O. 1980a. *On prosodic structure and its relation to syntactic structure*. Indiana University Linguistics Club.

Selkirk, Elisabeth O. 1980b. The role of prosodic categories in English word stress. *Linguistic Inquiry.* 11:563-605. [http://www.jstor.org/stable/4178179]

Selkirk, Elisabeth O. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, Mass.: MIT Press.

Selkirk, Elisabeth O. 1996. Sentence prosody: Intonation, stress and phrasing. In J. A. Goldsmith, ed. *The handbook of phonological theory*. London: Blackwell Publishers.

Selkirk, Elisabeth O. 2000. The interaction of constraints on prosodic phrasing. In Merle Horne, ed. *Prosody: Theory and experiment (studies presented to Gösta Bruce)*. 231-261. Dordrecht: Kluwer.

Selkirk, Elisabeth O. 2008. Bengali intonation revisited: An optimality theoretic analysis in which focus stress prominence drives focus phrasing. In Chungmin Lee, Matthew Kelly Gordon, and Daniel Büring, eds. *Topic and focus : Cross-linguistic perspectives on meaning and intonation*. 217-246. [New York]: Springer.

Steedman, Mark J. 1984. A generative grammar for jazz chord sequences. *Music Perception.* 2:52-77.

Steedman, Mark J. 1994. The well-tempered computer. *Philosophical Transactions: Physical Sciences and Engineering (Royal Society).* 349:115-131. [http://www.jstor.org/stable/54380]

Steedman, Mark J. 1996. The blues and the abstract truth: Music and mental models. In A. Garnham and J. Oakhill, ed. *Mental models in cognitive science*. 305-318. Mahwah, NJ:

Sundberg, Johan, and Björn Lindblom. 1976. Generative theories in language and music description. *Cognition.* 4:99-122. [doi:10.1016/0010-0277(76)90011-1]

Sundberg, Johan, and Björn Lindblom. 1991. Generative theories for describing musical structure. In Peter Howell, Robert West, and Ian Cross, eds. *Representing musical structure*. 245-272. London and San Diego: Academic Press.

Takano, Yuji. 2002. Surprising constituents. *Journal of East Asian Linguistics.* 11:243-301. [http://www.springerlink.com/index/P1H1425157248W76.pdf]

Tilsen, Sam, and Keith Johnson. 2008. Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America Express Letters.* [http://link.aip.org/link/?JASMAN/124/EL34/1]

Tojo, Satoshi, Yoshinori Oka, and Masafumi Nishida. 2006. Analysis of chord progression by HPSG. In V. Deved, ed. *Proceedings of the 24th IASTED international multi-conference: Artificial intelligence and applications*. 305-310. Innsbruck, Austria: ACTA Press.

Travis, Lisa. 1984. *Parameters and effects of word order variation*. Doctoral dissertation, Massachusetts Institute of Technology.

Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry.* 30:219-255. [http://www.jstor.org/stable/4179060]

Vergnaud, Jean Roger. 2006. Letter to Noam Chomsky and Howard Lasnik (1976). In Robert Freidin, and Howard Lasnik, eds. *Syntax: Critical concepts in linguistics*. Vol. 5. 21-34. London: Routledge.