

## CONTENTS

ELENI GREGOROMICHELAKI Grammar as action in language and music	1
---	---



---

# Grammar as action in language and music

ELENI GREGOROMICHELAKI

---

## 1 Introduction

Research in music and language cognition have followed similar paths in the last sixty years. Standard accounts of linguistic capacity, springing from early theories of transformational grammar and psycholinguistic information-processing models, investigate the *product* of processing, linguistic structured representations, conceptualised as internal, individualistic, idiosyncratic and encapsulated. Similar views were espoused in theories of music cognition where what was seen as underpinning a proper explanation of music and language capacities was “the supposition of specialized mental capacities, the belief that they could be studied rigorously by investigating the structure of their outputs” ([Lerdahl, 2009]:187). In these approaches, the structure of the output is seen as internalised within individual minds in the form of declarative knowledge that is employed during online processing.

However, it has recently been argued extensively that language and music share neural mechanisms, hence they might not each be specialised mental capacities at all. In this regard, it is notable that, whereas such models focus strongly on an account of music perception, processing in both language and music appears to be subsumed by the mechanisms involved in action. The common factor seems to be that all these domains involve time-linear sequential processing controlled by top-down expectations generated by potentially hierarchically-organised structures. If this is the case, *knowledge-how* [Ryle, 1949], rather than *knowledge-that*, appears then to be more explanatorily primary in the explication of such capacities. But, with language and music conceived as domain-specific cognitive modules (see e.g. [Jackendoff, 2002], [Jackendoff and Lerdahl, 2006]), theories of action would seem to have little relevance in accounting for their commonalities. Moreover, generative models where a “musical grammar,” abstractly associates strings of auditory events with musical structures [Lerdahl and Jackendoff, 1983] whereas a linguistic grammar is conceived as strictly separated from the conceptual-intentional and sensory-motor systems [Chomsky, 1993] do

not explain why there should be overlap of processing effects.

For language, the assumption that structural analysis of the output is of primary importance seems to be necessitated by a strict preoccupation with propositional semantic representations which are needed for the computation of inferences. This neglects other, perhaps more important, aspects of linguistic processing, namely, how language functions in coordinating human action, establishing social relations, achieving creative innovations and interacting with the environment. As a result, even in the domain of pragmatics, where language is viewed from an action perspective [Clark, 1996], mechanisms of agent coordination in conversation are assumed to be reducible to internal reasoning processes. These operate on the output of an internal, linguistic module represented in the mind of a given individual (I-language, [Chomsky, 1995]) and embed this in a series of metarepresentations, with the task of working out a predetermined speaker's intention/plan. To model this, computationally intractable inferential mechanisms, propositional attitude mindreading, strategic planning or game-theoretic deliberation are invoked, which contrast with the automaticity, fastness and efficiency that characterises online linguistic interaction. Similarly, in music, the composer's intention is implemented via their planned composition following some kind of music grammar [Schenker, 1935/1979] and this intention is then taken by some as paramount in evaluating/executing performances. Improvisation/collaboration in musical ensembles has also been seen as requiring a leader, a conductor or a soloist whose intentions direct and organise the group practice.

In contrast, recent models of joint action assign a central role to prediction in both action execution and action understanding, with subpersonal low-level online perception-action links being utilised to achieve the intersubjective understanding/coordination for which offline inferential models had previously been presumed to be needed. [Pickering and Garrod, 2013] apply these mechanisms to language production and comprehension in conversational dialogue for which there is a lot of evidence that predictive processes are crucially involved. In the domain of music, similar mechanisms have been argued to provide more realistic models of musical behaviour (see, e.g. [Pearce and Wiggins, 2006], [Pearce *et al.*, 2010] and the SAME model [Molnar-Szakacs and Overy, 2006]). Dynamic Syntax (DS, [Kempson *et al.*, 2001], [Cann *et al.*, 2005]), even more radically, takes an action-based architecture as constitutive of the grammar itself and presents a non-modular grammar model whose mechanisms are not assumed to be specifically linguistic. From this perspective, the close affinity between action, music and linguistic processing, contrasts sharply with the standard view of syntax as an abstract domain of knowledge, as assumed by orthodox

grammars. The core notion of DS is goal-directed incremental information growth/linearization following the time-linear flow of parsing/production without assuming an abstract representational level that imposes hierarchical structure over strings. Given this assumed close link between incremental/predictive processing and language structure, it emerges directly as a consequence that inherent features of the DS grammar architecture, employed to solve traditional grammatical puzzles, can also be shown to underlie many features of joint action, for example, language use in conversational dialogue. As a result, this chapter argues that low-level mechanisms like the grammar can in fact serve as the means for discovering one's own and others' emergent intentions during interaction without mind-reading as a prerequisite. Such a view extends naturally in the domain of musical collaboration and improvisation in so far as the claim is that automatic sensorimotor couplings provide the basis for group coordination online without the necessity of high-level cognitive processes regardless of modality and use.

## 2 Competence-performance in language and the nature of NL grammars

### 2.1 Syntax as the vehicle of propositional semantics

In the domain of linguistic capacity, the impact of the cognitive revolution of the fifties and sixties led to the development of the “Cartesian Linguistics” paradigm [Chomsky, 2009], with a focus on the private, internal, individual and rational aspects of human psychology. Under this conception, first of all, thought, a domain-general human capacity, and language, a domain-specific module, are separated in principle. The type of “thought” that language relates to is conceived as propositional in format, able to support structure-based deductive inferences (perhaps in some “language of thought”), with a referential semantics grounded by some kind of causal relation with the world. This is required under the adopted *computational theory of mind* presenting cognition as information-processing with an abstracted notion of *information* that relies on the opposition true/false (or 0/1) so that computational accounts of inferential capacities can be achieved. Following from this, natural language (NL) is not primarily seen as a means of communication but, instead, constitutes some arbitrary, formal expression of this type of thought/information. Given the internalist and structural character of the computational theory of mind approach, the expression of thought is conducted via the mediation of *syntax*, a level of analysis which explains how the apparently divergent surface structures of various languages map to universal thought structures. Accordingly, syntax is conceived as an idiosyncratic and autonomous level of mental states articulating how NLS

satisfy the constraints imposed on the expression of thought by systems that interface with external reality, namely, the conceptual-intentional (semantics) and the sensory-motor (phonology) [Chomsky, 1993]. One such constraint imposed by some is the presumed *compositionality* of semantic interpretation according to syntactic structure, i.e., the view that complex meanings are composed in lockstep with structured combinations of syntactic constituents.<sup>1</sup> This places a high burden on the syntactic component of a grammar since the surface NL structures are assumed to not reflect directly semantic compositionality, hence a host of semantically irrelevant, idiosyncratic transformational operations are needed to explain the relations between the interpretable structure (LF) and the sentence string [Heim and Kratzer, 1998].

## 2.2 The exclusion of dialogue data from competence theories

The fact that, from this point of view, linguistic grammars are concerned with the (optimal) organisation and expression of propositional referential thoughts entails particular commitments on the shape of such models. Firstly, NLS cannot be seen as cultural products or practices arising from or reflecting social interaction. Instead, as an expression of representational thought, NL reflects primarily the structure explicit conceptual thought imposes on the world. Meaning is restricted to the expression of such structures which are taken as what is transferred from individual to individual during “communication”. Accordingly, multiple linguistic phenomena reflecting issues of meaningful interaction among humans are assumed to require “syntactic” explanations in terms of the presumed underlying structuring of the string of words (the *sentence*) in terms of hierarchies involving (semantically-defined) *constituents*. This focus on phrase structure and its transformations to express some ideal ‘meaning’ naturally leads to the adoption of the formalism of *generative grammars* as a descriptive tool. However, because the matching between NL and thought structures is not one-to-one, such explanations are taken to involve an internal autonomous syntactic level [Newmeyer, 2010] immune from the influence of external constraints (*methodological solipsism*, [Fodor, 1980]). It then becomes a legitimate abstraction to ignore the mechanisms implementing the course of language perception and production, for example, the radical context-dependency of any semantic interpretations and the gradual availability of information as processing unfolds in a time-linear manner (*incrementality*). This is because it is assumed that the data structures involved (semantics, phonology) are already given independently at the in-

---

<sup>1</sup>Eventually this leads to identification of NL with thought, see e.g. [Hinzen, 2013], cf [Berwick and Chomsky, 2011] for related claims.

terfaces with non-linguistic components of cognition, hence syntax need only be concerned with defining the appropriate mappings, however complex or psychologically/phenomenologically unrealistic these might seem. As a result, from a methodological, but also, sometimes, a deep philosophical point of view, there is a circumscription of the relevant phenomena/data that grammars reflecting knowledge of language must account for. Knowledge of language is modelled through a *competence theory* characterising the syntactic-semantic-phonological representations and their interrelations. It is formulated as declarative knowledge that describes the structure of the eventual total product that results from the processing of whole sentences, in abstraction of either speaker, listener or the context of language use in general. Such presumed ‘mental’ grammars are shaped on the model of formal languages where a context-independent semantics directly relates to the syntactic structure. *Performance theories*, on the other hand, concern the interfaces with (representations of) external reality and hence may deal with issues like context-dependency or the constraints and possibilities afforded by how structure unfolds in real time. However, as a result, such theories bifurcate into independent components describing the use of grammar by either speakers (production) or listeners (parsing). Even non-“syntactocentric” theories like [Culicover and Jackendoff, 2005] espouse this view. The result of such exclusions and divisions is that the most natural arena of language use, everyday conversation, appears at best as incredibly complex from a performance point of view or completely irrelevant from the standpoint of a competence theory. This is because conversation data do not display the idealised sentence-to-proposition format required by a competence grammar. Instead, they consist of fragments (see e.g. turn 8 below) that are incrementally constructed and comprehended, and either then abandoned (turn 6, 7) or elaborated by the interlocutor (*split utterances*, see turns 3, 4, 5, 12, 14, 21), demonstrating that production and comprehension are tightly interlinked at the subsentential level:

- (1) 1. A: Instead of having *<name hidden>* *<unclear>* they had to come through the Dock Commission all of the men, they wanted so and so men for that boat, they used to come through to me.
2. B: Before that though, *<name hidden>* and *<name hidden>* [*<unclear>* had their own men ]
3. A: [ Had their own men
4. B: unload the boats?
5. A: unload the boats, yes. They *<unclear>*
6. B: They were employed directly by

7. A: That's right but they all came
8. B: *<name hidden>*?
9. A: They used to work say one week and have about a month off or go on the dole for a month.
10. B: So then what happened was, did the Dock Commission say you can't have your own men anymore?
11. A: That's right they had to go on a rota
12. B: Run by the Dock Commission?
13. A: Run by the Dock Commission. See the dockers then all got together and they said right so many men for that job, so many for that job and that didn't matter who they were, they had to *<unclear>* their job, all the way round the dock
14. B: Whether they wanted to go on that job or not?
15. A: Whether they want to go or not, they take their turn and the employer had to pay a percentage into the pool what those men earned, so when those men hadn't work at all they drew their money from the National Dock Labour Board.
16. B: Is this where the National Dock Labour Board came into existence?
17. A: That's how how they come into existence, yes *<name hidden>* he was a man what introduced that.
18. B: When was this?
19. A: Oh that's er, I would say about nineteen forty roughly *<clears throat>* I'd say about nineteen forty that came in, might have been before that.
20. B: Before that then if they were ill
21. A: They get nothing .
22. B: Could they not get any welfare benefit?
23. A: No [BNC, H5H: 89-113]

The reason for this is that meaning in language is not inherent in structured propositional thoughts exchanged between interlocutors. As studies have shown, the mechanisms that sustain interaction between individuals contribute in a crucial way to the development of meaningful exchanges. For example, [Schober and Clark, 1989] found that conversational partners



who were given the means of interacting with a speaker understood far better than overhearers who lacked this possibility even though the information conveyed through linguistic means was exactly the same. In addition, language use in conversation is highly dependent moment-to-moment during the interaction on integrating and combining inputs from several senses comprising non-verbal behaviors and features of the physical environment (*multi-modality*). For example, in face-to-face communication there is tight linguistic and embodied synchronization between speakers and listeners, with constant feedback loops jointly determining the course of the utterance as it unfolds via verbal and non-verbal signals [Goodwin, 1979; Goodwin, 1981; Goodwin, 1995] This is demonstrated in the excerpt below, adapted from [Bolden, 2003]. Besides the co-construction of utterances (*split utterances*, lines 2, 6, 10, 22), what are transcribed as pauses (indicated as durations in ⟨⟩ when of significant length) are in fact points where the language transcribed is highly indexical, composing online with actions, physical demonstrations of the apparatus discussed, gestures, gaze direction detection, body orientation etc:<sup>2</sup>

- (2) Simplified and adjusted text presentation of an excerpt from [Bolden, 2003]’s video-recorded data. The data come from a conversation between a lab technician (B) and a physicist (A) in which B describes the changes he plans to make to an antenna. B and A are standing next to each other examining an antenna prototype, which consists of a body with two parts, wheels and a belt. The prototype is in front of B and only B is handling it while explaining to A what the design involves now and what he plans to change.
1. B: I’ll put uh ⟨0.5⟩ idler ⟨*pause*⟩ wheel here. ⟨1.0⟩ And ⟨1.0⟩ the reason for that is (idler) wheel will come somewhere [ here.
  2. A: [ here ⟨0.6⟩
  3. B: It ⟨0.8⟩ gives us the tension aim and ⟨⟩ gives us ⟨1.0⟩ thuh ⟨1.2⟩ ninety degree or so wrap on[
  4. A: [Hm-mm Hm-mm ⟨0.4⟩
  5. B: ⟨*B is adjusting the antenna throughout*⟩ The problem with it like- ⟨0.4⟩ we had it ⟨0.8, *B is moving the top part of the antenna*⟩ now ⟨0.6⟩ is ⟨0.2⟩ you only have ⟨1.0, *B points*

<sup>2</sup>The online version of [Bolden, 2003] provides the audio-visual data. For reasons of space, here only parts pertaining to split-utterance processing are indicated.

- towards the belt of the antenna and turns his gaze towards A*  
 ⟨0.5⟩
6. A: **conjunction** .<sup>3</sup> ⟨0.5⟩ (punch)
7. B: ⟨*B is pointing towards the antenna and looking at A*  
 Three er- two [⟨unclear⟩
8. A: [⟨unclear⟩
9. B: ⟨0.5, *B briefly touches the middle wheel of the antenna*⟩ which  
 is ⟨*A nods affirmatively*⟩ ⟨*B makes circular movements pointing  
 to the wheel*⟩ When we try tuh move **it** real fast, ⟨0.5, *B rotates  
 the middle wheel of the antenna with his right hand*⟩ **it** ⟨0.4,  
*the belt slips and B shifts his gaze from the antenna towards A*⟩
10. A: Hh ⟨⟩ **slips** . ⟨0.4⟩
11. B: Yeah.
12. A: Mm
13. B: But if you get uh ninety degree wrap on it w[e should get it  
 pretty good
14. A: [Hm-mm ⟨2.2⟩
15. A: U[m
16. B: [So ⟨0.2⟩ it would give'em better ⟨0.4⟩ acceleration ⟨0.8⟩  
 deceleration cause right now ⟨0.2⟩ you can hear it when you  
 ⟨0.4⟩
17. A: Yeah
18. B: move it real fast, ⟨0.4⟩ the belt ⟨0.4⟩ slips [off.
19. A: [ ⟨uncertain⟩ ⟨2.4⟩
20. A: ⟨uncertain⟩ ⟨0.5⟩
21. B: And then **we'll also** ⟨0.6, *B moves the top part of the antenna  
 from the right to the left and A shifts his gaze slightly from the  
 antenna towards B*⟩
22. A: **move it** . ⟨0.8, *A is nodding*⟩
23. B: Yeah.

---

<sup>3</sup> *Conjunction* means “joining of parts” which refers to the juncture of the wheel and the belt.

Notice how the completions involve shifts of the participants' gaze towards each other.<sup>4</sup> Also, notice how meanings are incrementally affected due to the shifting interaction between language structure and the multi-modal context. For example, in line 9, because of the accompanying gestures and demonstration, the referent of the pronoun *it* (in bold) shifts without any problem within a single jointly-constructed sentence: firstly it refers to the wheel and then to the belt [Bolden, 2003]. What is demonstrated here more clearly for a pronoun like *it* is in fact the case for any lexical item and construction. Conversational participants follow each other's utterances and behaviours incrementally, perceiving and acting in the contextual situation where elements acquire variable meanings according to their temporal appearance in the string of words (and not just the sequential position of some overarching speech act as claimed by Conversation Analysis accounts).

From such evidence it can be concluded that linguistic competence is not exhausted by an account of coding and decoding decontextualised propositions conveyed via sentence structures. An account of competence needs to crucially involve an explication of the capacity of identifying, carrying out, and synchronizing social practices between individuals as well as their interactions with the physical environment. However, in competence/performance models, the modularity of the *language faculty* at either knowledge or mechanism level is a firmly rooted assumption with the result that the incremental subsentential/subpropositional interdependence and synchronisation of modalities and interlocutors cannot be accounted for in a unified framework.

In conclusion, standard accounts of linguistic capacity, springing from early theories of transformational grammar and psycholinguistic information-processing models, emphasize linguistic representations and processes conceptualised as internal, individualistic, idiosyncratic and encapsulated [Hauser *et al.*, 2002; Chomsky, 1995]. We now turn to examine similar assumptions in music analysis, the problems that are equally, and even more acutely, generated by such approaches and some suggestions for alternatives that might not only resolve the issues in the music domain but, in fact, provide guidance on how linguistic analysis should be conducted.

### 3 Competence theories in music

The focus on an ontologically prior conception of propositional semantics in linguistic research has justified a reification of NLS, with focus, not on the process, but rather the analysis of the (final) product derived during some idealised types of language use. This product is conceived as a complex hierarchical structure with transformational relations to other representations

---

<sup>4</sup>Use of gaze to manipulate the other interlocutor's involvement is a complex issue that we cannot go into here for reasons of space.

(needed to explain long-distance dependencies, quantifier scope etc.). Because of the assumed competence/performance distinction which relegates pragmatic or processing constraints to performance, such transformations came to be seen as having no semantic motivation, hence appearing *sui generis*, with the result that syntax was declared as an autonomous level of explanation. As a consequence, this kind of approach came to be seen as also applicable to music, as an explanation of its structural properties, despite its widely admitted non-referential, non-propositional nature,<sup>5</sup> which would seem to preclude any direct transfer of the semantically-motivated hierarchical structurings attributed to NLS.

### 3.1 An account of musical structure: Generative Theory of Tonal Music

A competence theory of a particular style, Western tonal music, is developed in [Lerdahl and Jackendoff, 1983] (Generative Theory of Tonal Music, GTTM henceforth, see also [Lerdahl, 1988]). Its purpose is to characterise the structures underlying sequences of notes, the musical “phrase structures”, that constitute a piece of Western tonal music according to a listener’s intuition. As Lerdahl and Jackendoff state, instead of describing the listener’s real-time mental processes, they are concerned only with “a grammar of tonal music”, the final state of an idealised experienced listener’s understanding. Focusing on composed musical scores, the abstract structures postulated in this model constitute, in their view, what a listener assigns to a piece when they “understand” it as music (rather than hearing it as a sequence of musical notes, the *musical surface*). Following standard assumptions, as articulated by [Marr, 1982], regarding the separation of levels of analysis, computational, algorithmic, implementational, this is justified because, in their view, it would be fruitless to theorize about mental processing before understanding the product that this processing derives (i.e. their analysis stands at the Marrian computational level). Hence, they define an internalised mental grammar for music, building on the Schenkerian tradition in music theory [Schenker, 1935/1979], and, as a methodological hypothesis they take it that product/process aspects of music cognition can be cleanly separated (for an opposing view at the level of general cognitive modelling, arguing against Marr’s distinctions, see e.g. [Sun *et al.*, 2005]). The grammar aims to express those components of knowledge of music that, in their view, are hierarchical in nature with the result that other dimensions like timbre, dynamics and motivic-thematic development, but also the idiosyncratic expressive variation introduced by performers

<sup>5</sup>See e.g. [Kivy, 2002], except in cases like programme music etc, see e.g. [Patel, 2008]:6.3.1 for other cases

which is what makes music a joint achievement, are not part of the model. Because of the lack of a referential propositional semantics to justify the proposed “phrase structures” and semantically-derived notions like *headedness* of phrases,<sup>6</sup> Gestalt psychological principles and intuitively derived assignments of points of tension and resolution (i.e., instability vs. stability; openness or closure) are invoked.

Accordingly, in GTTM, a musical piece is seen as the total product described by a musical score rather than an auditory sequence of events that evolve through time. Despite GTTM’s immense contribution in raising issues highly pertinent to musical analysis and cognition in general, as in NL competence theories, time-linearity is treated as having few if any consequences for the structural analysis. Gestalt principles like ‘grouping’ that equally apply to vision are also applied to the analysis of musical pieces. Because the unifying view of integrative online processing is missing, the result is that, as in Jackendoff’s current linguistic model (see e.g. [Jackendoff, 2002]), multiple independent levels of description are needed. The musical surface is associated with four kinds of autonomous levels of analysis with feeding relations among them. As regards rhythm, *grouping structure* defines hierarchically-organised constituents that partition a piece into motives, phrases, and sections whereas *metrical structure* associates a piece with a hierarchical grid of strong and weak beats. Both these analyses feed into the construction of two pitch-related structures: *Time-span reduction* encodes relative structural importance through assignments of dominating and elaborative roles to each note. Building on the tradition of [Schenker, 1935/1979], *prolongation reduction* follows time-span reduction and defines a structure of ‘tension’ and ‘resolution’ in harmonic terms with elaboration or contrast as the main determinants. Each level of musical structure is characterised by a set of *well-formedness rules* specifying all the possible structural analyses, and *preference rules* designate which structure is more likely to be assigned to the particular piece under analysis. Transformational rules, of much less importance than in NL grammars, apply distortions to the hierarchical structures ensuing from the well-formedness rules.

Even more paradoxically for music than language, where issues of propositional semantics obscure the temporal dimension of analysis, this type of approach falls under the characterisation of “architectonicism” [J. Levinson 1997] in that it treats music perception statically, more like the perception of a work of architecture or painting (the paradigm of ‘auditory scene anal-

---

<sup>6</sup>A harmonic “semantics” as attributing points of rest and their preparation to chord sequences is invoked by [Steedman, 1984] and developed into a mental-model type harmonic implication semantics in [Steedman, 1996]; pragmatic principles of “coherence” are advocated in [Patel, 2008]: 6.3.2).

ysis' [Bregman, 1990] adopted by Jackendoff and others also alludes to this general type of focus on structural perception). What is crucially lacking is an appreciation of how the sequential nature of perception/production (*incrementality*) and the context-dependency of processing (*multi-modality*) affect the nature of the phenomena examined. [Jackendoff, 1991] delineates a parsing model of how the musical surface can be assigned the structures postulated by GTTM but crucially, and naturally for such a model with a clear separation of the competence theory from performance, he does not mention how music production, composition or any other type of musical use like improvisation, interpretation, performance, etc. would contribute to explanatory aspects of the model. Moreover, as pointed out in the implementation of [Widmer, 1995], several significant linear connections between events are missed by GTTM's strictly hierarchical analysis.

### 3.2 Temporal accounts of music perception

Although also framed as a competence theory, Narmour's *implication-realisation model*, (IR henceforth, Narmour [1992; 1977]), building on work of Meyer [Meyer, 1956, 1973], eschews GTTM's multiple-level hierarchical structuring and is explicitly concerned with the perception of music in time focusing specifically on the continuity of melodic relations and conflating GTTM's independent levels of analysis. IR relies on note-to-note relations characterised as more or less similar or different, which provides the basis for identifying patterns of realising or thwarting listeners' expectations. Importantly, because the time-linear direction is taken into account, as in NL processing, the model admits of temporary ambiguity/underspecification which is resolved as further structure becomes available, an aspect completely missed in GTTM. From a psychological point of view, "bottom-up", hard-wired unconscious principles and involuntary mechanisms are assumed to implement first-order music perception. The kind of structures appearing in GTTM's analyses are characterised as involving issues of "style", learned associations and schemata, imposed top-down to influence a piece's perception according to the listener's expertise. Thus, despite its initially promising integrative and time-linear perspective, in the end, IR imposes an even stricter Fodorian modular approach on music perception separating the (potentially) hierarchical style-"grammar" from hard-wired innate musical perceptual principles.

A more extreme view towards the significance of incremental perception is taken by J. Levinson's *concatenationism*, based on [Gurney, 1880], although the theory is presented at a more intuitive, impressionistic level [J. Levinson 1997]. These models seek to describe the phenomenological experience of a musically untrained listener as they attend to a piece of music.

This experience is strictly focused in the present, in the music segment currently being processed and its immediate connection to the segments immediately preceding or following. However, as in Narmour’s IR, the listener’s experience of the musical present may be affected by memories of the musical past or expectations for the musical future. These memories and expectations are typically not precise or conscious, but rather vague and subliminal. In contrast, listeners with higher expertise might engage in more “intellectual” listening, involving conscious assignment of structures to the piece, perhaps structures like the ones described by GTTM or the “folk-psychological” ones [Cross, 1998; Wiggins *et al.*, 2010] described by music theory. Unlike what has been claimed for NL, where the derivation of a propositional structure is elevated as the principal aim of processing given its role in inferencing, in J. Levinson’s view, the parts of a successful musical composition, linked through “cogency of succession” principles (e.g. chord progression or voice-leading) are valuable in themselves, affording the primary enjoyment of aural involvement with a piece. The large-scale form, as described by Shenkerian analyses, J. Levinson argues, provides only a secondary or ‘parasitic’ kind of enjoyment.

### 3.3 Perceptual music theories and music use

Narmour’s IR and the GTTM models are not mutually exclusive (see e.g. [Widmer, 1995] for a combination). However, in both, the focus is on individual minds with internal perceptual and computational machinery engaged in the manipulation of internal representations (as dictated by the computational theory of mind hypothesis). And, in the case of music, this comes without the justification provided by the a priori referential semantics that have been employed in NL analyses. Even J. Levinson’s analysis, despite its focus on one processing aspect, the temporal organisation of the musical surface, ignores other crucial aspects of “music use”, crucially, production, either composition or performance, group improvisation and participation, arguably the most widespread uses of music across societies [Cross, 2012]. As in theories of NL, these aspects of music behaviour, are supposed to be issues for theories of performance to deal with. We turn now to examine how the separation of phenomena in competence/performance terms that seek to preserve the independence of competence theories, both in music and NL, present a distorted picture of the phenomena they are supposed to explain.

## 4 Performance in language and music

Various aspects of musical activities and behaviours, when considered in performance models incorporating competence theories, are subsumed under

the view of communication that presupposes some version of the *code model* for NLS. Messages (i.e. scores) are generated internally, encoded through a grammar specifying the code and subsequently decoded by the listener, perhaps through the mediation of a passive performer, using the same code grammar. This kind of picture makes much less sense for music than it does for NL given that the nature of music’s “meaning” indicates that the “code” does not encrypt preformed intentional propositional messages with truth-conditional semantics. However, this has not prevented the advancement of prescriptivist attitudes [Cook, 1999] towards music performance or composition by the proponents of competence models that are based on how (they think) perception works. [Lerdahl, 1992] requires conformity between the ‘compositional grammar’ and the ‘listening grammar’ (but cf [Lerdahl, 1997] distinguishing between “natural” and “artificial” compositional grammars), whereas [Narmour, 1988] provides criteria for judging correct or incorrect performances (reminiscent of the grammatical/ungrammatical distinction in language).

#### 4.1 Issues of correctness/“grammaticality” in language and music and the delimitation of evidence

Formal theories of NL, assumed to reflect competence, have long abandoned the mere ambition of separating grammatical from ungrammatical sentences at the string level.<sup>7</sup> Instead, the proposed analyses are justified on the basis of semantic intuitions as regards constituency, headedness and other interpretational phenomena like movement, anaphora etc. So these analyses presuppose an already independently identifiable propositional semantic structure to which syntax, via various intermediate representations, maps the phonological representation. In the domain of music such propositional structure is absent. From the present point of view, this is one of the reasons why a notion of “grammaticality” cannot be applied to musical pieces in the same way as some presume it applies to language. But even in NLS, because performance-related explanations are in principle excluded, context-independent assignment of grammaticality characterisations frequently breaks down because not even the semantics can be used to define context-independent propositional structures. In case after case, the effect of various aspects of the context have to be “grammaticalised” as part of the code, leading to artificial distinctions and characterisations. For example, the ban of linear explanations for (an already restricted set of) coreference constraints (*Binding Theory*:  $\text{John}_i$  dislikes himself<sub>*i*</sub>/\*him<sub>*i*</sub>;

---

<sup>7</sup>This statement is intended to go beyond the weak/strong generative capacity distinction. The claim is that strong generative capacity has a semantic justification. Distributional criteria do not produce sufficiently unambiguous results.



\*Himself<sub>i</sub> likes John<sub>i</sub>) led to elaborate hierarchical structure explanations and ad hoc syntactic mechanisms whose complexity ensued to the abandonment of the whole Binding Theory grammar component as either a purely conceptual-intentional interface issue [Chomsky, 1993] or a purely pragmatic issue [Levinson, 2000].

Similarly, and more acutely, in music. Even if we accept GTTM's preference rules as legitimate mechanisms to be incorporated in a competence model (for justification see [Jackendoff and Lerdahl, 2006]), there are other cases where the abstractions necessitated by use of a generative grammar become untenable. For example, melodic and rhythmic/metrical structure requirements might clash with the assumed harmonic structure principles. As a result, this is one case where transformational rules of significant expressive power are employed in a competence theory with independent levels of description such as GTTM and, even more widely, in performance models employing such "neutral" grammars (see e.g. [Widmer, 1995]). In the same vein, to justify lack of applicability of the abstracted structural constraints, it is often stated that a work of art almost necessarily will break the established rules in some way [Steedman, 1984], hence a competence theory should not be concerned with such "exceptions", relegated to a theory of performance, but rather with the core rules that the analyst's intuitions deem as appropriate.

However, we can take an alternative view of the significance of this observation. Instead of taking it as reflecting some deep differentiation between music, or art in general, and NL, which has come to light because of the attempt to apply a generative theory to both domains, this observation rather points to the fact that the definition of "grammaticality" in generative theories of NL, defined on the basis of whole sentences and derived in abstraction of processing mechanisms and context, should equally be abandoned. The assumption that the surface order of words, or the musical surface, is necessarily underpinned by the hierarchical structuring of the string in terms of the assignment of elements to headed syntactic categories, is an assumption that derives from semantic intuitions that ground NL processing to propositional information exchange. The idealisation of NL semantics as concerning the construction of logical forms, independently of performance factors that, like music, involve embodiment, emotion and expression conflicts with a realistic account of NL/music uses. For example, the incremental nature of the linguistic/musical experience which requires and allows constant flexibility and adaptability to the surrounding context leads such static models to ignore the most naturally occurring linguistic uses, as in (1)-(2) earlier, excluding them as ungrammatical/irrelevant. In parallel to (1)-(2), the musical experience has similar characteristics, most

evident in group improvisation settings, which is a domain completely unaccounted for in the competence models mentioned earlier which focus strictly on individualistic perception. However, group improvisation, like conversation, is grounded by the possibility for joint action resulting in the musical surface displaying co-construction of musical events:

- (3) But you see what happens is, a lot of times when you get into a musical conversation, one person in the group will state an idea or the beginning of an idea and another person will complete the idea or their interpretation of the same idea, how they hear it. So the conversation happens in fragments and comes from different parts, different voices. (Ralph Peterson cited in [Monson, 1996]: 78 wherein musical examples of such “conversations”)

Moreover, as in conversation, the products of joint action that result from such exchanges are emergent [Sawyer, 2003] and not accounted for by investigating each individual’s particular contribution:

- (4) This is a story about me and three other musicians. Led by Ade Knowles, we were rehearsing a piece based on Ghanaian musical principles. Each of us had a bell with two or three heads on it –the bells were of Ghanaian manufacture. Ade assigned three of us simple interlocking rhythms to play and then improvised over the interlocking parts. Once the music got going, melodies would emerge that no one was playing. The successive tones one heard as a melody came first from one bell, then another and another. No one person was playing that melody; it arose from cohesions in the shifting pattern of tones played by the ensemble. ... Occasionally, something quite remarkable would happen. When we were really locked together in animated playing, we could hear relatively high-pitched tones that no one was playing. ... The tones were distinct, but not ones that any of us appeared to be playing. (from [Benzon, 2001]: 23-24)

In our view then, the potential for joint action during performance has irreducible effects pertaining to the explanation of the derived product whose structure competence theories aim to account for. In joint action the aim is not the transmission of propositional messages but rather the coordination of participants to act together. Moreover, during performances, music frequently occurs along with language, theater, movement, dance or ritual and religious themes so that the emergent product does not simply relate autonomously to the auditory domain. These are not issues that can be addressed in a competence model, however, their effects affect the structure of the product these theories analyse. Given this lack of explanatory coverage, are there any other justifying assumptions for the adoption of competence models?

#### **4.2 Hierarchical structuring and recursion: primitives or products of action-oriented cognition?**

Competence theories take the view that the structure of the final product can be described independently, irrespective of how it emerged during

performance, and that the discoveries that result from such investigations are valuable because they reveal crucial aspects of human psychology. For example, even without propositional semantic intuitions, there is one remaining argument justifying the use of generative grammars for both NL and music. This is the assumption of infinite generation of surface auditory or linguistic patterns from a finite set of rules by employing recursive embedding.<sup>8</sup> [Hauser *et al.*, 2002] assume that such *recursion* is a unique cognitive characteristic of NL, uncovered via the assumption of the internal cognitive capacity that generative grammars model. However, in contrast, S. Levinson [2013] argues that recursive embedding is not a feature of the *syntax* of some languages, hence it can't be a universal NL-syntactic characteristic. [Steedman, 2002] and Jackendoff [2011; 2009] point out that all cognitive capacities of any complexity, including music, display recursion. Jackendoff then goes on to state that NL is unique primarily in that combinatorial recursion in the communicative signal, i.e., the sentence string, maps into combinatorial recursion in the message conveyed, i.e., the semantic structure.

It is exactly this observation that shows why the “separationist” methodological strategy employed by competence models, instead of explaining, actually leads to proliferation of explanatory levels, without an independent remit, and unjustifiable circumscription of phenomena. In this case, the observation is that most of the NL syntactic constituency criteria and transformational operations are given a semantic justification (with the rest being pragmatic/processing factors that have been unjustifiably “grammaticalised”). We are then led to the question whether Jackendoff's assumed NL syntactic combinatorial structure is just an epiphenomenon attributed to the recursively-hierarchical organisation of certain types of thought that traditional AI models favour, i.e., thought that reflects (conscious, voluntary) propositional reasoning and inferencing. In music, recent proposals attempting to unify NL and music grammars, postulate a common level of syntactic hierarchical structure that is only semantically-interpretable in the case of NL [Katz and Pesetsky, 2009; Tsoulas, 2010]. However, this again introduces an unjustified redundancy, given the semantic justification of such hierarchical structures for NLs. In music, where no such ontologically prior propositional semantic structure applies, analysts' intuitions are not that firmly in favour of necessary underlying hierarchical structure.<sup>9</sup> The first-order time-linear coherence of the music stimulus, as explicated by Narmour's IR and J. Levinson's models, can be taken as the primary

<sup>8</sup>See also [Jackendoff, 2011] for the notion of recursion intended and cf [Lobina, 2011] for clarification of the notion of recursion in general.

<sup>9</sup>For some issues leading to a potentially contrary view see [Rohrmeier *et al.*, to appear]

dimension of analysis, expressed perhaps via the assumption of a simple time-linear rule-based or Markovian model (see also [Pearce and Wiggins, 2006], [Pearce *et al.*, 2010] cf [Giblin, 2008]). Culturally-determined learned hierarchical structuring constituting a particular “style” can then be seen as developing as a result of familiarity/expertise with musical forms (via the adoption of musical *schemata* or heuristics, see e.g. [Krumhansl and Castellano, 1983] and more pertinently [Butler, 1989], [Brown *et al.*, 1994]). For NLs, it has been argued that the NL syntactic categories and the apparent hierarchical structuring of the string can be eschewed and attributed solely to the semantic structure derived during processing ([Kempson *et al.*, 2001], [Cann *et al.*, 2005]; see also [Steedman, 2000]). From this point of view, hierarchical structuring and recursion are then plausible features of (some of) the semantic/conceptual structure utilised both by NLs and (types of) non-linguistic cognition as well as, perhaps, culturally-defined style schemata in music. As such, these assumptions make sense but they do not adequately characterise the primary mechanisms explaining the function of cognition in these domains. Instead, a procedural model, a “grammar” based on the dynamics of processing, is fundamentally required in order to explicate how structured representations are derived and used online, serving as the generators of incremental, context-dependent expectations for both NL and music (see e.g. [Kempson *et al.*, 2001], [Pearce and Wiggins, 2006], [Pearce *et al.*, 2010], Kempson & Orwin [this volume], Orwin [this volume]).

From this point of view, hierarchical structuring and recursion are not features uniquely related to either NL or music. Instead, NL and music can be seen as forms of action (production/generation) and action perception (comprehension/parsing). Hierarchical structuring and recursion can then be taken as consequences of the fact that both NL and music inherit the (joint-) action oriented nature of cognition expressed as skillful performance, knowing-how, across various tasks which can be combined and accounted for in a single non-encapsulated framework. However, traditional cognitive and AI theories have investigated the structure of human action in a static way and solely by reference to the contents and organization of individual minds. We turn now to the problems associated with such an approach and what alternatives are available.

## 5 Action and performance in music and language

### 5.1 Static action and perception grammars

Jackendoff [2011; 2007] employs his own version of generative grammar to suggest that actions are recursively structured, involve *headed* constituents and even variable binding and long-distance dependencies in ways quite analogous to what, he assumes, characterise NL syntax. Following standard

planning models, the compositional organisation that he assumes underlies complex actions, like making coffee, involves the compilation of a hierarchy of subactions stored in long-term memory. The basic constituent is the “elaboration” of an action into a *head* (or the ‘core’ action, e.g. setting the machine to produce the coffee), an optional *preparation* subaction (e.g. fill the machine with water and coffee), and an optional *coda* rounding off the action or reestablishing the status quo (e.g. put away the coffee jar). Each of these constituents may itself be elaborated into preparation, head, and coda. These elaboration structures are highly reminiscent of the GTTM model’s prolongation structures as well as Jackendoff’s version of NL syntax.

However, despite these apparent similarities with a presumed activity-neutral grammar of action, GTTM (as well as Narmour’s IR) are models dealing solely with declarative knowledge pertaining to music perception. Aspects of production, especially in the type of music that these models address, raise a whole host of issues that these models, even more readily than linguistic ones, are prepared to relegate to performance as highly intractable (see earlier (3)-(4)). As a result they cannot be reconciled with recent evidence that perception and action, two domains that have been considered separately in traditional cognitive analyses, do not operate independently, especially when cognitive mechanisms, rather than declarative knowledge structures, are examined. Looked at from this point of view, issues that arise in music/NL production, joint performances, conversation, composition, planning etc have to be considered as primary factors in the explanation of the cognitive infrastructure that results in products, which, in idealised cases, display the hierarchical structuring assumed in static grammars of action. But the modularised view of cognition that up to now has characterised standard models does not allow a straightforward integration.

## 5.2 The “cognitive sandwich” view of cognition

Instead, competence/performance models, both in the domain of music and NL, presuppose what [Hurley, 2008] has characterised as the “cognitive sandwich” view. According to this view, the mind is structured at three levels: perception and action are seen as separate from each other and peripheral; cognition, the locus of propositional thought, planning, and executive control, stands in between as the filling. Further, this view postulates that low-level perception and action involve a series of independent modules which are separate from the higher processes of cognition that provide the only interface among them (Fodor’s *central systems*). Cognition/thought display a series of related properties like compositionality, systematicity, productivity, binding, etc., which are explained solely in terms of processes involving combinatorial syntactic structure ([Fodor and Pylyshyn,

1988] a.o.). This is the basis of the classical computational-theory-of-mind approach and it leads directly to competence-performance models for NL. Since modular performance domains are independent, but nevertheless requiring some common properties especially at their interfaces with central thought/cognition, a neutral module “grammar”, specifying what is needed for interfacing with the cognitive filling, has to be assumed.

NL performance is in turn explicated through some version of the code model that accounts for the transfer of “meanings”, conceived as propositions at the cognitive level, from one individual mind to another. When such a view was transferred to music, it led to describing music capacity in a modular fashion too. Accordingly, Narmour’s IL adopts a strict modular architecture, Lerdahl states that GTTM provides a set of hypotheses about the structure of a mental music module [Lerdahl, 1997]: 392 whereas [Jackendoff, 2009] argues that there is a need to posit a “narrow musical capacity” since the properties of music do not all follow from other more general cognitive principles (see also [Pinker, 1997]). Even recent conjectures about NL and music following the generative paradigm find only very superficial similarities between them, namely that Merge applies recursively to create headed hierarchies, which, in the case of NL, can be interpreted as a separate step via a propositional semantics [Katz and Pesetsky, 2009; Tsoulas, 2010]. Yet these modularity assumptions have been disputed. [Pearce and Wiggins, 2006] dispute Narmour’s distinction between bottom-up and top-down influences on expectation. Instead, in their model, both the bottom-up principles and style influences ensue as the results of general-purpose learning mechanisms acquiring descriptions of regularities through exposure to music, which are then expressed as patterns of expectations.<sup>10</sup> Dynamic Syntax [Kempson *et al.*, 2001; Cann *et al.*, 2005; Gregoromichelaki *et al.*, 2011] and [Pickering and Garrod, 2013] dispute modularity claims based on NL evidence.

### 5.3 Common mechanisms for action, language and music processing

Further, in contrast to the modular view, recent research has shown that both music and NL processing involve the same mechanisms. Patel [2003; 2008] presented experimental evidence for this claim and also identified overlapping brain areas, including Broca’s area, as involved in processing in both domains (see also [Hagoort, 2005], [Abrams *et al.*, 2011], [Maess *et al.*, 2001], [Sammler *et al.*, 2009]). Examining performance in both domains, it has been shown that the unexpected violation of regularities in NL or music affects processing in the same way displayed by the elicitation of similar

<sup>10</sup>However they allow that such learning can lead to domain-specific representations.

electric brain potentials and the observation of interference effects when the violations are presented simultaneously [Koelsch *et al.*, 2005; Fedorenko *et al.*, 2009; Slevc *et al.*, 2009; Steinbeis and Koelsch, 2008]. In addition, it has been argued that syntactic processing deficits affect both domains in parallel [Fazio *et al.*, 2009; Sammler *et al.*, 2009; Grodzinsky, 2000; Patel *et al.*, 2008], while processing in both domains can be improved by training in only one of them [Jentschke and Koelsch, 2009; Marin, 2009]. On the basis of such evidence, it has been claimed that, as regards mechanisms,<sup>11</sup> syntactic processing in both NL and music share resources. The common link seems to be the interrelation of both domains with the motor system and therefore action. [Fadiga *et al.*, 2009] and [Pulvermüller and Fadiga, 2010] among others, review research indicating that the same brain regions are involved not only in NL processing and music but also in action execution and observation (see also [Pulvermüller, 1999], [Rizzolatti and Craighero, 2004], [Rizzolatti and Craighero, 2007]).

#### 5.4 Generative grammars explain action perception and execution?

Processing and executing sequential steps, in a goal-directed manner, seems to be the common factor between language, music and action. In addition, in contrast to the “cognitive sandwich” point of view, direct mappings and *common coding* (see [Hommel *et al.*, 2001]) for production/perception processes have been argued to underlie all three domains. So from a theoretical point of view, all skilled acts pose the problem of appropriate sequential ordering that has to be resolved during execution. As with Jackendoff’s [2011; 2009] attempt seen earlier, the problem can be approached by the method of constructing a generative grammar defining abstractly the hierarchical structuring of the task and justifying the sequential order in hierarchical terms. [Pastra and Aloimonos, 2012] also define a minimalist generative grammar for action including long-distance dependencies and binding. However, by conceiving this structuring in abstraction from performance, (and despite the “processing-friendly” aspects of the ‘unification’ operation as argued by [Jackendoff, 2011]), separate parsers/generators have to be defined for various uses. This misses the essence of the common coding perspective in that it conceives of the mind again as consisting of independent input-output, perception-action modules that require mediation from a general cognitive store structuring and storing knowledge in a declara-

---

<sup>11</sup>The fact that the experimental evidence shows dissociations between knowledge representations for music and NL [Patel, 2008] might preclude models like [Pearce and Wiggins, 2012], [Pearce and Wiggins, 2006] from being able to account for the demonstrated commonalities in processing in that they seem to conflate structured representation and processing. The same applies for the minimalist grammar of [Phillips, 1996].

tive manner. As in music and NL tasks, as exemplified earlier in (1)-(4), the radical context-dependency of executing/comprehending actions online renders unrealistic the long-term storage of hierarchically structured complex action sequences assumed to provide top-down control over execution or attribution. Instead, incremental parsing/generation of local expectations/goals is a better strategy that is flexible and adaptable to current contextual conditions whether these involve interaction with the physical environment or social coordination (see e.g., for NL: [Kempson *et al.*, 2001], [Gregoromichelaki *et al.*, 2011]; for music: Pearce and Wiggins [2006; 2012], Kempson and Orwin [this volume]).

In conclusion, in all three domains, NL, music and action, models based on classical planning architectures (e.g. [Bratman, 1987], [McDermott, 1978]) face the problem of how the overarching structure which constitutes the basis of their explanation (e.g. a Shenkerian prolongation structure in music, a propositional intention in NL and a ‘core’ action in planning) can ever be incrementally arrived at and, once established, how it either is recognised or guides processing. Especially in the domain of social joint action where comprehension and execution need to interweave tightly, the phenomenon of action sharing in most widespread uses (see (1)-(4) earlier) poses difficulties for such models.

### 5.5 Intention recognition explains joint action?

Unlike NL, until the overwhelming recent developments in Western societies of recorded music and non-participatory listening, music has always been associated with motor action and, consequently, the social action perspective has been taken in music studies. According to [Cross, 2012], ethnomusicology research suggests that music should primarily be taken as a medium for human interaction enabling and constituted by social processes. [Small, 1998] argues for the introduction of the term “musicking” and a similar proposal, “linguaging” [Linell, 2009], has been made for NL focusing on the emergence of the NL system from the practices that constitute NL uses. The grounding of fundamental NL structures to action is argued for by S. [Levinson, 2013]. However, he goes on to argue that mapping of sounds and multimodal signals onto speech acts are recognized through Gricean intention recognition, presupposing advanced theory of mind capabilities, audience design and constituting a component of the mind with its own dedicated neural circuitry [De Ruiter *et al.*, 2010]. Levinson then attempts to extrapolate these assumptions to music.

From the present point of view, this analogy is misleading. Levinson seems to hold on to the language-as-product paradigm that presupposes standard information-processing analyses springing from early cognitivist



competence theories emphasising linguistic representations based on propositional thinking and focusing on individual cognitive processes. Faced with the evidence of joint action and radical context-dependency as in (1)-(4) earlier, the code-model is enriched with computationally intractable inferential mechanisms, propositional attitude mindreading, strategic planning or game-theoretic deliberation which are postulated to account for joint activity mediated through NLS. This strategy generates puzzles like the *mutual knowledge paradox* [Clark and Marshall, 1981], according to which, interlocutors have to compute an infinite series of beliefs in finite time which contrasts with the automaticity, fastness and efficiency that characterises online interaction. The intractability of such solutions then, in turn, provides arguments that enhance the competence/performance separation as well as modularity and “cognitive sandwich” assumptions.

Given the lack of propositional semantics for music (or its “floating intentionality” [Cross, 2003]), this approach makes much less sense for this domain, hence if there are any commonalities between music and NL, as the neurobiological evidence seems to suggest, necessary Gricean intention recognition in the NL domain becomes a burden. There have been attempts to reconceptualise the classical (neo-) Gricean accounts of communication in terms of implicit subpersonal and interpersonal processes, sometimes even rejecting the Belief-Desire-Intention (BDI) model of explanation while attempting to maintain that inferential mental state ascription is the primary basis for communication (see e.g., [Sperber and Wilson, 1995], [De Ruiter *et al.*, 2010], [De Ruiter *et al.*, 2007], [Davies and Stone, 1995] a.o.). However, from the present point of view, such attempts risk introducing unnecessary conceptual confusion in two respects. Firstly, the view that attribution of mental states is the sine-qua-non for communication is taken as axiomatic, rather than a position to be defended (see also [de Bruin *et al.*, 2011]) thus ignoring a range of alternatives to be explored (see e.g. [Ginzburg, 2012], ch. 7; [Gregoromichelaki *et al.*, 2013b], [Mills, 2011], [Piwek, 2011], [Mills and Gregoromichelaki, 2010], Mills [this volume]). Secondly, as a consequence of this stance, even when behaviours, situations or domains like music are tackled that are not properly explained through the necessary attribution of folk-psychological abilities (e.g. lack of “theory of mind” evidence in animals/infants/autistic patients, context-dependency/vagueness of speech act content, collaborative emergence of structures and intentions in conversation/music), researchers still seek to postulate something weaker as a substitute, exploiting then the Marrian computational/algorithmic distinction to treat such constructs as the mechanisms enabling “intention recognition”. What is missed here is that attribution of propositional attitude mindreading is only justified under the assumption that the agents

understand, employ and engage with the complex causal structures that the logic of such states requires (see e.g., [Davidson, 1980]; for further explication see [Apperly, 2011], Ch. 5; [Bermúdez, 2003]). Especially for Gricean intentions, this should involve multiple levels of metarepresentation. More pertinently for our purposes here, from the point of view of standard psychological and computational models where communication is conceptualised as crucially involving Gricean propositional-attitude mindreading, interspersed within low-level processing steps, NL conversational interaction appears to be very complex (see e.g. [Poesio and Rieser, 2010]) for an admirably thorough illustration of this complexity in accounting for a single type of split-utterances). This is because in conversation, as can be seen earlier in (1)-(2), interlocutors must be modelled as able to deal with fragmentary utterances, which are produced/comprehended incrementally so that they can be abandoned or modified, before a sentence/proposition has been constructed. In addition, such fragments both compose with, and are interpretationally dependent on, the physical environment and the other interlocutor’s subsentential feedback actions. So interlocutors must be able to switch rapidly between production and comprehension, perform processing at both levels simultaneously [Pickering and Garrod, 2004], and develop plans/intentions on the fly.

In music, and art in general, the paramount importance of deriving the creator’s intention is a long-discussed and disputed issue (see e.g. [Wimsatt and Beardsley, 1946], [Dipert, 1980], [Kivy, 1998]), involving the ideal of *Werktreue* [Goehr, 1992], and frequently resulting on focus being placed on the composer’s intended interpretation rather than the performer’s contribution. However, when focus is placed on musical creativity and improvisation, research on collaborative group performance [Sawyer, 2003] suggests that Gricean metarepresentation is not the most appropriate explanatory mechanism. Thus, we propose that, instead of conceptualising musical exchanges as ‘communication’, which echoes the legacy of the ‘code model’, we should focus our attention on musical *coordination*. In ensemble performances experienced musicians seem to coordinate their actions based on familiarity and development of joint routines (see e.g. Mills [this volume]). In jazz group improvisation settings, it has been argued that no goal, especially not a propositional one, can be defined external to the improvisation process and the ensuing performance is an emergent result not reducible to explanation in terms of individual minds [Sawyer, 2008]. In contrast to Gricean metarepresentation, [Leman, 2008] replaces classical versions of propositional intention recognition (“cerebral intentionality”) with *corporeal intentionality* conceptualised as an emerging effect of the coupling of action and perception. Even in rehearsed ensemble performance, the focus is shifting from

plan/intention recognition to embodied skills and low-level mechanisms like ‘behavioural resonance’ [Leman, 2008], entrainment [Clayton *et al.*, 2005], anticipation, auditory imagery, and movement simulation [Keller, 2008; Keller *et al.*, 2007; Repp and Knoblich, 2004].

In our view, there are lessons to be learned for NL research from these developments. Taking the focus away from propositional thought exchanges and informational uses of NLS, a new perspective of thinking of NL in action terms emerges from applying models of the motor system to NL processing. [Pickering and Garrod, 2013] and Dynamic Syntax at the NL domain are two of those. Unsurprisingly, these models investigate language use in conversational dialogue where traditional accounts fail to offer intuitive modelling.

## 6 Dialogue within action-based frameworks

As already noted, it is in turning to dialogue modelling that competence-based frameworks face their stiffest challenge, for the explanation of the systematicity of dialogue turns on bringing together ‘language-as-product’ and ‘language-as-action’ perspectives. Two NL models which take up this integration challenge as a design feature, [Pickering and Garrod, 2013] (P&G henceforth) and Dynamic Syntax, will be examined below from that perspective. In the domain of music, the SAME model [Molnar-Szakacs and Overy, 2006; Overy and Molnar-Szakacs, 2009] and [Pearce and Wiggins, 2012], [Pearce *et al.*, 2010] seem also compatible with this point of view.

### 6.1 Common coding, simulation and coordination

*Pickering and Garrod [2013]*

The model presented by P&G develops the basis of a psychological account of coordination that promises to provide a compromise between the ‘language-as-product’ and ‘language-as-action’ paradigms in a way that reconciles realistic fast processing in dialogue with the interpersonal and sub-personal mechanisms that support fluent intersubjectivity. Standard modular accounts of NL separate production and comprehension by postulating an intermediate cognitive level of integration (i.e. “the cognitive sandwich” perspective), a view that is incompatible both with the demands of communication and with extensive data P&G present indicating that production and comprehension are tightly interwoven at a very fine-grained level. P&G’s conclusions are also supported by cases like those shown in (1)-(2) earlier, where interlocutors clarify, repair and extend each other’s utterances, even in the middle of an emergent clause (*split-utterances*) switching fluently among planning, comprehension, production and integration of contextual cross-modal inputs. In order to solve the puzzle of rapid and

fluent NL-based interaction, P&G propose to conceptualize NL processing in terms analogous to recent accounts of action attribution and execution. In the light of current evidence regarding common coding between perception and action (e.g., [Bargh and Chartrand, 1999], [Sebanz *et al.*, 2006]), neurocomputational accounts have been developed that make use of the notion of ‘internal models’ (e.g., [Grush, 2004], [Wolpert *et al.*, 2003] see also [Hurley, 2008]). On these views, during execution of goal-directed actions, it is more efficient to derive and use a predictive (*forward*) *model* of the expected dynamics rather than simply waiting to react on the basis of actual reafferent feedback. Accordingly, during execution, an ‘efference copy’ of the motor command is created causing the forward action model to generate the predicted act and its consequences, which are then compared with the actual feedback for adjustment and learning purposes. Similarly, during perception, an *inverse model* (plus the context) can be used to covertly imitate the actor and predict their subsequent movements thus either leading to overt imitation or achieving goal-understanding as well as coordination in joint action cases. In these accounts of goal-directed action, a central role is assigned to *prediction* in both action execution and action understanding, with subpersonal low-level online perception-action links being utilised to achieve the intersubjective understanding/coordination for which offline inferential models had previously been presumed to be needed. P&G apply these mechanisms to NL production and comprehension for which there is a lot of evidence that they crucially involve predictive processes (e.g., comprehension: [Levy, 2008]; production: [Pickering and Garrod, 2007; Florian Jaeger, 2010]). According to P&G, speakers use forward models to predict their upcoming utterances thus adjusting their output accordingly (*audience design* phenomena could be taken as based on such a mechanism, but see also [Gann and Barr, 2012], [Horton and Gerrig, 2005]). Listeners covertly imitate speakers through use of inverse models which, through learned associations and the shared current context, provide the background for understanding the speaker’s “intention” in uttering the current input. They then use forward models based on their own potential next motor command to predict what speakers are likely to say next (this constitutes the “simulation route” to comprehension).

### *The SAME model*

A similar architecture at the neural level underlies the Shared Affective Motor Experience (SAME) model [Molnar-Szakacs and Overy, 2006; Overy and Molnar-Szakacs, 2009] which investigates music processing that results in shared affective states. This model is based on evidence that the human mirror neuron system (MNS) shows sensitivity to auditory stimuli related to actions [Aziz-Zadeh *et al.*, 2004; Buccino *et al.*, 2004]. Following suggestions

in the NL and action domains, it proposes a common neural substrate for music, NL and motor functions based on the coupling of action and perception afforded by the MNS. This constitutes an automatic and unconscious simulation mechanism where the same neural resources are utilised both to perform one's own actions and represent and understand the actions of others [Gallese, 2003]. Since motor acts are coded in the MNS as belonging to an action sequence, this mechanism also enables prediction of observed action goals thus facilitating the coordination of individuals as regards intentional and emotional states without cognitive mediation. (For similar ideas see also [Leman, 2008]).

## **6.2 Dynamic Syntax: fine-grained incrementality and predictivity in dialogue and the role of grammar**

### *Eschewing multiple representation levels*

Despite the radical nature of their model, from the present point of view, P&G maintain a conservative stance as regards the online progress of interaction, rehearsing standard assumptions about how NL processing is executed. Similarly to standard models like [Jackendoff, 2011] for NL, as well as GTTM in music, they assume that linguistic information has to be organised hierarchically and represented at different levels between message and articulation: (at least) semantics, syntax, and phonology. These levels are ordered “higher” to “lower,” so that a message (including speech act characterisations) generates a semantic representation, semantics evokes a syntactic representation, this in turn maps to phonology, and from phonology to speech sounds. Thus, a production process goes from message to sound via each of these levels (message  $\mapsto$  semantics  $\mapsto$  syntax  $\mapsto$  phonology  $\mapsto$  sound) whereas a comprehension process goes from sound to message in the opposite direction. Given the forward model that speakers and listeners both use to predict what is likely to come next, this means that producing utterances involves not only production processes but also comprehension processes; similarly, comprehending utterances involves comprehension processes but also incorporates production processes. Crucially, reflecting the relationship between the linguistic levels, the production command is taken to constitute the message that the speaker wishes to convey, including information about intended speech act, pragmatic context, and a nonlinguistic situation model, which is then mapped to the representational levels assumed at the action execution phase. It is this assumption that, in our view, causes problems for the P&G account when applied to a wide range of dialogue data. The reason is that, as in other performance models that aim to incorporate competence theories, in such an analysis, incrementality is only simulated rather than being part of the architecture. So we look

instead at a more fine-grained incremental account, Dynamic Syntax (DS), where time-linearity is an architectural feature of the grammar *ab initio*.

Work within the action-based Dynamic Syntax (DS) model makes similar assumptions as P&G regarding the tight interlinking of NL perception/production in that both speakers and listeners have to perform mirrored context-dependent actions in order to integrate or produce NL strings incrementally (see [Cann *et al.*, 2005], Kempson & Orwin [this volume]), but, perhaps, in diverse contextual environments since the cognitive circumstances of each agent might be distinct. Given the fine-grained incremental DS architecture, assumed to model the grammar of NLs, efficiency dictates that processing is not strictly bottom-up but instead guided by predictions ('goals'). These are expectations dictated by either the integration of current NL input or generated as general top-down computational goals. As speakers and listeners simulate the actions of each other, the fulfillment of these goals is due at each incremental step, subsententially, for both parser/generator and can be satisfied by either, on the basis of the other interlocutor's input or by recourse to the processor's own resources and context. As no structure is ever assumed to be derived for the sentence string, no whole string grammaticality considerations arise and hence processable fragments and split utterances are directly licensed and, in fact, a natural consequence of such a fine-grained bidirectional incremental system. For this reason, from an interpretational point of view, Dynamic Syntax predicts a much wider range of split-utterance types than the P&G model with its standard message-syntax-semantics articulation.

The P&G model is perhaps able to cope with the type of split-utterances termed *collaborative completions* as in (6) and (5):

- (5) Helen: When I left you at the tube earlier, I went home and  
found my boyfriend...  
James: in bed with another woman. Shit! [Sliding Doors]
- (6) Joe: We were having an automobile discussion ....  
Henry: discussing the psychological motives for  
Mel: drag racing in the streets. [Sacks 1992: 144-145]

However, it is very much less compatible with the many other types of continuations in conversation. As (7)-(9) show, such completions by no means need to be what the original speaker actually had in mind, so an account of their generation does not need to involve prediction at the message or semantic levels:

- (7) Helen: I, I'm sure you're not a nutcase or a psycho or anything, it's just that, um I'm not, I'm not that good at, um you know, um...  
James: Constructing sentences? [Sliding Doors]
- (8) Helen: I love this bridge. My great grandfather helped to build it. I often come and... stand on it when I want to, um...  
James: Build a bridge? I'm sorry [from Sliding Doors]
- (9) Connie: Clarence, I am looking for you! Where are you? I want to talk to you! Clarence?  
(*Connie bangs hard on cupboard's door where Clarence is hiding*)  
Clarence: Ah, Connie, splendid! Erm... Heard you calling. Wasn't able to find you, so I thought, what a capital idea to...  
Connie: Fling the servants' shoes around? [from Blandings:  
Pig-hoo-o-o-ey! BBC2 14/1/13]

Like the emergent phenomena in musical group improvisation that we saw earlier reported in (4) (see also [Sawyer, 2008]), in (7)-(9), the string of words ('sentence') that the completion yields is not at all what either participant would have planned from the beginning. The same goes for the message (or semantic representation). In such cases and many others (see [Gregoromichelaki *et al.*, 2013a]), contra to S. Levinson's [2013] assumption that mindreading is necessarily involved in NL action coordination, there is no reason to suggest here that, before interrupting, the listener first figured out the original speaker's plan, then derived the expected continuation, then rejected it, then figured out a new plan which resulted in an alternative continuation which he/she then produced, while the original speaker went through the reverse process in order to comprehend and integrate this continuation.

Such data then cast doubt on the Gricean assumption, a residue of the code model, that in all successful acts of communication, the speaker must have in mind some definitive propositional content which they intend to convey to their hearer, whose task, conversely, is to succeed in grasping that particular content. Some variant of this assumption underpins many current pragmatic theories (see e.g. [Bach and Harnish, 1979], [Sperber and Wilson, 1995], [Levinson, 2000]). But this assumption is just an artifact of the NL models assumed where a [sentence (syntax)  $\Leftrightarrow$  proposition (semantics)] mapping for each utterance is required. This is on the basis of the employment of (a) competence generative grammars that need to evaluate whole sentences as (un)grammatical and (b) classical cognitive inferential models that rely on propositional deductive reasoning. However, in actual

use, speakers do not have to have fully-formed propositional intentions in order to start speaking. Instead, the sequential nature of the conversational structure (see e.g. [Schegloff, 2007]) as well as, in general, subpersonal mechanisms like those assumed in the SAME model discussed earlier in section 6.1, crucially now incorporated in the grammar formalism, provide an adequate background for accounting for the emergence of joint structures and negotiated meanings.

*Eschewing necessary intention-recognition*

Unlike standard assumptions as in P&G and Jackendoff’s models, where an intended speech act has to be generated to achieve the appropriate multi-level mappings, given the sequential context provided by the conversation, multiple speech acts can be performed by use of a single grammatical construction shared across turns between interlocutors:

- (10) A: Go away  
 B: and if don’t  $\langle$ conditional antecedent  $\Rightarrow$  Continuation; Question $\rangle$   
 A: I’ll smash your face  
 $\langle$ conditional consequent  $\Rightarrow$  Continuation; Reply; Threat etc. $\rangle$   
 [natural data]
- (11) Freddie (who fancies the boss’s daughter): I didn’t know  
 you were ...  
 Mike (who goes out with boss’ daughter):  
 banging the boss’ daughter?  $\langle$ Completion/Clarification $\rangle$   
 [Cemetery Junction]

Notice that these are not just cases of “one action being the vehicle for another” (or indirect speech acts) as identified by S. [Levinson, 2012] and [Schegloff, 2007]. Here multiple actions are performed during the unfolding of a single propositional unit. Therefore, at an appropriate sequential environment, co-construction can be employed for the performance of speech acts without first establishing propositional contents. Moreover, based on the fact that syntax and interpretation are both conceptualised as a single action system, [Gregoromichelaki *et al.*, 2013a] argue that actions in dialogue can be accomplished just by establishing “syntactic conditional relevances”, i.e., exploiting the grammatical dependencies themselves to induce a response by the listener (*grammar-induced speech acts*). In the following, incomplete syntactic dependencies can be initiated by a speaker inviting the listener to fulfill them thus forming a question-answer pair during the derivation of a single proposition:



- (12) A: Thank you mister ...  
 B: Smith, Tremuel [natural data]
- (13) A: Shall we go to the cinema or ...  
 B: let's stay at home [natural data]
- (14) A: And you're leaving at ...  
 B: 3.00 o'clock
- (15) Man: and this is Ida  
 Joanna: and she was found?  
 Man: she was found by a woman at Cheltenham. [Catwoman]
- (16) A: And they ignored the conspirators who were ...  
 B: Geoff Hoon and Patricia Hewitt [radio 4, Today programme, 06/01/10 ]
- (17) Jim: The Holy Spirit is one who  $\langle \rangle$  gives us?  
 Unknown: Strength.  
 Jim: Strength. Yes, indeed.  $\langle \rangle$ The Holy Spirit is one who gives us?  
 $\langle \rangle$   
 Unknown: Comfort. [BNC HDD: 277-282]
- (18) George: Cos they  $\langle$ unclear $\rangle$ they used to come in here for water and bunkers you see.  
 Anon 1: Water and?  
 George: Bunkers, coal, they all coal furnace you see, ... [BNC, H5H: 59-61]

There is no reason to suppose here that the speaker had a fully-formed propositional message to convey before they started production, in fact these formats exactly contradict various assumed [speech act  $\leftrightarrow$  syntax] mappings. Moreover, in some contexts, invited completions of another's utterance have been argued to exploit the vagueness/covertness/negotiability of the speech act involved to avoid overt/intrusive elicitation of information:

- (19) (Lana = client; Ralph = therapist)  
 Ralph: Your sponsor before ...  
 Lana: was a woman  
 Ralph: Yeah.  
 Lana: But I only called her every three months.

Ralph: And your so your sobriety now, in AA [(is)]

Lana: [is] at a year.

Ralph: A year. Well, I'm not perhaps the expert in this case at all. However, I must admit that you're still young in (.) sobriety and I think that maybe still working with a woman for a while might be

Lana: Yeah

Ralph: in your best interest.  
[from Ferrara 1992]

Here the therapist uses an invited completion in a way that gives the patient the opportunity to assign it the force of question or not and hence to reveal or not as much information as she is willing to reveal.

As argued in [Kempson *et al.*, 2009], [Gregoromichelaki *et al.*, 2011], Kempson & Orwin [this volume], what is essential in accounting for all these data, along with “disfluencies” which abound in actual conversation (see earlier examples (1)-(2)) is an incremental grammar that models the parallel course and common mechanisms of parsing/production at an appropriate subsentential/subpropositional level. Along with other researchers, we have suggested that intentions/plans should not be seen as causal factors driving coordination but, instead, as discursive constructs that are employed by participants, as part of a (meta-)language regarding the coordination process itself, when participants need to conceptualise their own and others' performance for purposes of explicit deliberation or accountability when trouble arises. Empirical evidence for this approach come from studies showing that, in task-oriented dialogue experiments, explicit negotiation is neither a preferential nor an effective means of coordination [Garrod and Anderson, 1987]. If it occurs at all, it usually happens after participants have already developed some familiarity with the task. Further more specific evidence has been provided by experiments probing participants' awareness of even their own intentions in early and late stages of task-oriented dialogue leading to expert performance (see e.g. [Mills and Gregoromichelaki, 2010], [Mills, 2011], [Mills, 2013], Mills [this volume]). It has been shown that as participants develop more and more expertise in the task, awareness of plans/intentions emerges and can then be utilised as a means of coordination when trouble ensues (see also [Suchman, 2007]).

*An action-based conception of grammar and the achievement of coordination*

For these reasons, in our view, the production/comprehension of fragments and split-utterances in conversation cannot be taken to causally rely on the determination of a pre-planned speaker-intended speech-act. Indeed,

in our view, preplanned joint intentionality is uncommon in dialogue: to the contrary, joint intentionality has to develop through engagement with the task, via subpersonal, subconceptual mechanisms, therefore, as argued in [Sawyer, 2008] for musical improvisation, it is emergent rather than constitutive of joint action. One such mechanism that the participants share *ab initio* is a set of processing routines and practices, in our view, the “grammar”, that can ground further coordination via the predictive goal-directed processing and mirroring that it imposes. From this point of view, the important observation that comes from split-utterance data is that their licensing crucially employs this grammar. As shown in (Kempson & Orwin [this volume]) in more detail, and earlier in (1)-(2) and (5)-(19) the dependencies binding each part of a split-utterance span over the entire range of syntactic and semantic dependencies, and are observable in all languages [Howes *et al.*, 2011; Purver *et al.*, 2009; Kempson *et al.*, 2012]. Given that such dependencies are licensed grammar-internally, a grammar formalism has to be able to handle the combination of such fragments if it is to meet minimal conditions of adequacy. However, these data are highly problematic for all standard frameworks, given the commitment to models of NL knowledge (competence grammars) licensing such dependencies over sentence-strings independent of any performance realisation.

In contrast, Dynamic Syntax (DS) assumes an action-based formalism for the characterisation of the combinatorial properties of NL. In effect, on this view, the grammar emerges from the sedimentation of motor mechanisms originally evolved to control/represent the hierarchical structure of instrumental action (for a similar view of how “syntax” emerged, see also [Gallese, 2007], section 8; [Hurley, 2008]; [Pulvermüller and Fadiga, 2010]). Thus, in parallel to assumptions in the P&G model, but more radically transferred within the grammar itself, the DS combinatorial mechanisms employ an architecture similar to those assumed in the control of the hierarchies that emerge in the analysis of goal-directed actions. But since these mechanisms constitute a relatively fixed and stable architecture that can be employed rapidly, reliably and automatically, there is no need to assume the necessary employment of forward/inverse models whose usual function is in the service of learning and adjustment. Instead, predictivity/goal-directedness is built right inside the operation of the grammar for efficiency and control purposes. That is, the grammar design includes a top-down element that provides the source for the generation of predictions (which can further be simulated in a forward model but need not necessarily be so); and the coupling of parser/generator is intrinsically modelled as a form of covert imitation and prediction through the employment of identical mech-

anisms in a shared context. Such predictions guide lexical retrieval at a subpropositional level, for both speaker and listener in parallel, irrespective of what role they realise currently. It is this more basic mechanism (at a similarly low-level as the “association route” in the P&G model) that participants exploit in the generation of split-utterances in order to steer the conversation towards their own goals without necessarily having to consider the current speakers’ intended messages (which they also can do employing mechanisms as those outlined in the P&G, the SAME models and [Pearce and Wiggins, 2006]). Under this view, participants can progress via an associative route, guided by the goals generated by the grammar and, on this basis, negotiate derivative constructs like intentions and strategies overtly at the social level (“externalised inference”, see also [Pickering and Garrod, 2004]). For music, such a model can be supplemented with an account of learning as advocated in [Pearce and Wiggins, 2006], [Pearce *et al.*, 2010] to provide the requisite flexibility and context-dependent adaptation. This approach has the advantage that it does not fall under the criticism leveled against Meyer’s [1956] and Narmour’s IR models that musical expectations (predictions) and their resolution cannot support a theory of musical affect generation because, contrary to fact, familiarity should end up obliterating emotional involvement [Jackendoff, 1991]. In a potential music model based on DS assumptions, rather than P&G’s, predictions and their resolution are what drives low-level processing and, if such predictions are assumed to generate affect, this will be the result of processing regardless of the familiarity or not of the piece currently processed. This is the same as in linguistic processing where a “meaning” will be derived however familiar or predictable an utterance is to the listener. Affect, what some consider as the “meaning” of music, and we would argue, in part, language as well, is then invoked as the result of automatic, unconscious subconceptual processing rather than by trying to divine some kind of intention, musical or linguistic.

Seen from this perspective, the P&G model represents a significant advance within the language-as-action paradigm in providing a mechanistic non-inferential account for action understanding and production in dialogue. However, we suggest that in maintaining several aspects of the language-as-product tradition, namely, standard multi-level mappings between sound and meanings, it does not go far enough in extending the action-based architecture, hence it also does not provide a suitably domain-general system in order to incorporate an account of music processing.

## 7 Conclusion

Reconceptualising the grammar along the lines suggested by DS promises to solve another problem having to do with the relevance of neuroscience evi-

dence for models of NL competence. Linguists have long disputed the compatibility of current theories of brain function with (competence) theories of syntactic structure (see e.g. [Jackendoff, 2002]). Because no alternative to standard competence models has been conceived, such claims take it for granted that the alleged abstract nature of syntactic structure, as an intermediate level between sound and meaning, conflicts with the requisite direct matching between perceptual linguistic information and corresponding motor plans that recent neuroscience models advocate. Especially for the kind of evidence that P&G cite, regarding the close affinity between action and NL processing, current neuroscience results pointing in the same direction, and the commonalities between NL processing and music, the view of NL syntax as an abstract domain of declarative knowledge, as assumed by standard grammars, constitutes the biggest stumbling block for further progress (as also noted by [Patel, 2008], section 5.4.3). This standard view of syntax as an abstract intermediary has led to specific claims that this immunity to brain evidence is due to the very nature of syntactic phenomena that are, it is claimed, not amenable to time-linear sequential explanations ([Tettamanti and Moro, 2012]; cf. [Pulvermüller, 2010]). According to this standard view, syntactic explanations rely on complex hierarchical structures that become hidden to the bodily senses due to their linearisation into strings of words. Hence, it is claimed, this inaccessibility to perceptual systems implies that syntactic processing must rely on different capacities than those involved in matching perceptual linguistic information onto corresponding motor plans as assumed in the P&G and the SAME models.

However, from the DS perspective presented earlier, there is an alternative action-based view of “syntax” which makes it directly commensurate with architectures like the P&G and SAME models as well as with currently proposed neurobiological mechanisms mediating action understanding/execution. All “idiosyncratic” syntactic phenomena identified by [Tettamanti and Moro, 2012], as well as phenomena like long-distance dependencies and binding assumed to hold in both NL and music (see e.g. [Jackendoff, 2011], [Thompson-Schill *et al.*, 2013]) are modelled in DS in action-based procedural terms, i.e. as involving knowledge-how rather than declarative knowledge of multi-level mappings. Hence the assumption of common processing mechanisms for both NL and music becomes a real possibility, since both are seen primarily as involving processes rather than representational constructs. As such, in our view, both NL and musical ability cannot be separated from socio-cognitive mechanisms of joint action and situated processing. The focus of music models like GTTM, Narmour’s IR and more recent ones like [Katz and Pesetsky, 2009], [Tsoulas, 2010] on perception and representation ignores the fact that the basic function of both cognition and

perception is in the service of controlling action.<sup>12</sup> As shown within DS, by focussing on mechanisms that underpin coordination between interacting individuals, rather than the “communication” of propositional messages from one individual mind to another, a unified account of both music and language can begin to emerge.

## BIBLIOGRAPHY

- [Abrams *et al.*, 2011] Daniel A Abrams, Anjali Bhatara, Srikanth Ryali, Evan Balaban, Daniel J Levitin, and Vinod Menon. Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex*, 21(7):1507–1518, 2011.
- [Apperly, 2011] Ian Apperly. *Mindreaders: The Cognitive Basis of theory of Mind*. Psychology Press, 2011.
- [Aziz-Zadeh *et al.*, 2004] Lisa Aziz-Zadeh, Marco Iacoboni, Eran Zaidel, Stephen Wilson, and John Mazziotta. Left hemisphere motor facilitation in response to manual action sounds. *European Journal of Neuroscience*, 19(9):2609–2612, 2004.
- [Bach and Harnish, 1979] Kent Bach and Robert M Harnish. *Linguistic communication and speech acts*, volume 4. MIT press Cambridge, MA, 1979.
- [Bargh and Chartrand, 1999] John A Bargh and Tanya L Chartrand. The unbearable automaticity of being. *American psychologist*, 54(7):462, 1999.
- [Benzon, 2001] William Benzon. *Beethoven’s anvil: music in mind and culture*. Oxford University Press, 2001.
- [Bermúdez, 2003] José Luis Bermúdez. The domain of folk psychology. *Royal institute of Philosophy Supplement*, pages 25–48, 2003.
- [Berwick and Chomsky, 2011] Robert Berwick and Noam Chomsky. The biolinguistic program: The current state of its evolution and development. In Anna Maria Di Sciullo and Cedric Boeckx, editors, *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*, pages 19–41. Oxford: Oxford University Press, 2011.
- [Bolden, 2003] Galina B Bolden. Multiple modalities in collaborative turn sequences. *Gesture*, 3(2):187–212, 2003.
- [Bratman, 1987] Michael E. Bratman. *Intentions, Plans, and Practical Reason*. CSLI Publications, 1987.
- [Bregman, 1990] Albert S Bregman. *Auditory scene analysis: the perceptual organization of sound*. MIT press, 1990.
- [Brown *et al.*, 1994] Helen Brown, David Butler, and Mari Riess Jones. Musical and temporal influences on key discovery. *Music Perception*, pages 371–407, 1994.
- [Buccino *et al.*, 2004] Giovanni Buccino, Stefan Vogt, Afra Ritzl, Gereon R Fink, Karl Zilles, Hans-Joachim Freund, and Giacomo Rizzolatti. Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. *Neuron*, 42(2):323–334, 2004.
- [Butler, 1989] David Butler. Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception*, pages 219–241, 1989.
- [Cann *et al.*, 2005] Ronnie Cann, Ruth Kempson, and Lutz Marten. *The Dynamics of Language*. Elsevier, Oxford, 2005.
- [Chomsky, 1993] Noam Chomsky. *A Minimalist Program for Linguistic Theory*. MIT Press, 1993.

---

<sup>12</sup>To the extent that [Pearce and Wiggins, 2006], [Pearce *et al.*, 2010] do not account for the direct pairing of perception/production, the same criticism might apply (but see [Pearce and Wiggins, 2012]: 643–44, for an initial perspective on these issues).

- [Chomsky, 1995] Noam Chomsky. *An Essay on Minimalism*. MIT Press, 1995.
- [Chomsky, 2009] Noam Chomsky. *Cartesian linguistics: A chapter in the history of rationalist thought*. Cambridge University Press, 2009.
- [Clark and Marshall, 1981] Herbert H. Clark and C. R. Marshall. Definite reference and mutual knowledge. In *Elements of discourse understanding*. Cambridge: Cambridge University Press, 1981.
- [Clark, 1996] Herbert H. Clark. *Using Language*. Cambridge University Press, 1996.
- [Clayton *et al.*, 2005] Martin Clayton, Rebecca Sager, and Udo Will. In time with the music: The concept of entrainment and its significance for ethnomusicology. In *European Meetings in Ethnomusicology*, volume 11, pages 3–142, 2005.
- [Cook, 1999] Nicholas Cook. Analysing performance and performing analysis. In N. Cook, & M. Everist editors, *Rethinking Music*, pages 239–61. Oxford University Press Oxford and New York, 1999.
- [Cross, 1998] Ian Cross. Music analysis and music perception. *Music Analysis*, 17(1):3–20, 1998.
- [Cross, 2003] Ian Cross. Music and evolution: Consequences and causes. *Contemporary music review*, 22(3):79–89, 2003.
- [Cross, 2012] Ian Cross. Cognitive science and the cultural nature of music. *Topics in Cognitive Science*, 4(4):668–677, 2012.
- [Culicover and Jackendoff, 2005] Peter W. Culicover and Ray Jackendoff. *Simple Syntax*. Oxford University Press, 2005.
- [Davidson, 1980] Donald Davidson. Toward a unified theory of meaning and action. *Grazer Philosophische Studien*, 11:1–12, 1980.
- [Davies and Stone, 1995] Martin Davies and Tony Stone. *Folk psychology: the theory of mind debate*. Blackwell, 1995.
- [de Bruin *et al.*, 2011] Leon de Bruin, Derek Strijbos, and Marc Slors. Early social cognition: Alternatives to implicit mindreading. *Review of Philosophy and Psychology*, 2(3):499–517, 2011.
- [De Ruiter *et al.*, 2007] Jan Peter De Ruiter, M Noordzij, Sarah Newman-Norlund, Peter Hagoort, and Ivan Toni. On the origin of intentions. *Attention & Performance XXII*, pages 593–610, 2007.
- [De Ruiter *et al.*, 2010] Jan Peter De Ruiter, Matthijs L Noordzij, Sarah Newman-Norlund, Roger Newman-Norlund, Peter Hagoort, Stephen C Levinson, and Ivan Toni. Exploring the cognitive infrastructure of communication. *Interaction Studies*, 11(1):51–77, 2010.
- [Dipert, 1980] Randall R Dipert. The composer’s intentions: An examination of their relevance for performance. *The Musical Quarterly*, 66(2):205–218, 1980.
- [Fadiga *et al.*, 2009] Luciano Fadiga, Laila Craighero, and Alessandro D’Ausilio. Broca’s area in language, action, and music. *Annals of the New York Academy of Sciences*, 1169(1):448–458, 2009.
- [Fazio *et al.*, 2009] Patrik Fazio, Anna Cantagallo, Laila Craighero, Alessandro D’Ausilio, Alice C Roy, Thierry Pozzo, Ferdinando Calzolari, Enrico Granieri, and Luciano Fadiga. Encoding of human action in Broca’s area. *Brain*, 132(7):1980–1988, 2009.
- [Fedorenko *et al.*, 2009] Evelina Fedorenko, Aniruddh Patel, Daniel Casasanto, Jonathan Winawer, and Edward Gibson. Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*, 37(1):1–9, 2009.
- [Ferrara, 1992] Kathleen Ferrara. The interactive achievement of a sentence: Joint productions in therapeutic discourse. *Discourse Processes*, 15(2):207–228, 1992.
- [Florian Jaeger, 2010] T Florian Jaeger. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1):23–62, 2010.
- [Fodor and Pylyshyn, 1988] Jerry A Fodor and Zenon W Pylyshyn. Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1):3–71, 1988.
- [Fodor, 1980] Jerry A Fodor. Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, 3(01):63–73, 1980.

- [Gallese, 2003] Vittorio Gallese. The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology*, 36(4):171–180, 2003.
- [Gallese, 2007] Vittorio Gallese. Before and below ‘theory of mind’: embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):659–669, 2007.
- [Gann and Barr, 2012] Timothy M Gann and Dale J Barr. Speaking from experience: Audience design as expert performance. *Language and Cognitive Processes*: 1–23, 2012.
- [Garrod and Anderson, 1987] Simon Garrod and Anne Anderson. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27:181–218, 1987.
- [Giblin, 2008] Iain Giblin. *Music and the generative enterprise*. PhD thesis, Doctoral dissertation, University of New South Wales, 2008.
- [Ginzburg, 2012] Jonathan Ginzburg. *The Interactive Stance: Meaning for Conversation*. Oxford University Press, 2012.
- [Goehr, 1992] Lydia Goehr. *The Imaginary Museum of Musical Works: An Essay in the Philosophy of Music*. Oxford University Press, 1992.
- [Goodwin, 1979] Charles Goodwin. The interactive construction of a sentence in natural conversation. In G. Psathas, editor, *Everyday Language: Studies in Ethnomethodology*, pages 97–121. Irvington Publishers, New York, 1979.
- [Goodwin, 1981] Charles Goodwin. *Conversational organization: interaction between speakers and hearers*. Academic Press, New York, 1981.
- [Goodwin, 1995] Charles Goodwin. Co-constructing meaning in conversations with an aphasic man. *Research on Language and Social Interaction*, 28(3):233–260, 1995.
- [Gregoromichelaki et al., 2011] Eleni Gregoromichelaki, Ruth Kempson, Matthew Purver, Greg J. Mills, Ronnie Cann, Wilfried Meyer-Viol, and Pat G. T. Healey. Incrementality and intention-recognition in utterance processing. *Dialogue and Discourse*, 2(1):199–233, 2011.
- [Gregoromichelaki et al., 2013a] Eleni Gregoromichelaki, Ronnie Cann, and Ruth Kempson. On coordination in dialogue: subsentential talk and its implications. In Laurence Goldstein, editor, *On Brevity*. Oxford University Press, 2013.
- [Gregoromichelaki et al., 2013b] Eleni Gregoromichelaki, Ruth Kempson, Christine Howes, and Arash Eshghi. On making syntax dynamic: The challenge of compound utterances and the architecture of the grammar. In Ipke Wachsmuth, Jan de Ruiter, Petra Jaacks, and Stefan Kopp, editors, *Alignment in Communication: Towards a New Theory of Communication*. John Benjamins, 2013.
- [Grodzinsky, 2000] Yosef Grodzinsky. The neurology of syntax: Language use without broca’s area. *Behavioral and Brain Sciences*, 23(01):1–21, 2000.
- [Grush, 2004] Rick Grush. The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27(3):377–396, 2004.
- [Gurney, 1880] Edmund Gurney. *The Power of Sound*. Smith, Elder, 1880.
- [Hagoort, 2005] Peter Hagoort. On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, 9(9):416–423, 2005.
- [Hauser et al., 2002] Marc D Hauser, Noam Chomsky, and W Tecumseh Fitch. The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598):1569–1579, 2002.
- [Heim and Kratzer, 1998] Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Blackwell Oxford, 1998.
- [Hinzen, 2013] Wolfram Hinzen. Narrow syntax and the language of thought. *Philosophical Psychology*, 26(1):1–23, 2013.
- [Hommel et al., 2001] Bernhard Hommel, Jochen Müsseler, Gisa Aschersleben, and Wolfgang Prinz. The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24(05):849–878, 2001.



- [Horton and Gerrig, 2005] William S Horton and Richard J Gerrig. Conversational common ground and memory processes in language production. *Discourse Processes*, 40(1):1–35, 2005.
- [Howes *et al.*, 2011] Christine Howes, Matthew Purver, Patrick G. T. Healey, Gregory J. Mills, and Eleni Gregoromichelaki. On incrementality in dialogue: Evidence from compound contributions. *Dialogue and Discourse*, 2(1):279–311, 2011.
- [Hurley, 2008] Susan Hurley. The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioural and Brain Sciences*, 31:1–58, 2008.
- [Jackendoff and Lerdahl, 2006] Ray Jackendoff and Fred Lerdahl. The capacity for music: What is it, and what’s special about it? *Cognition*, 100(1):33–72, 2006.
- [Jackendoff, 1991] Ray Jackendoff. Musical parsing and musical affect. *Music Perception*, pages 199–229, 1991.
- [Jackendoff, 2002] Ray Jackendoff. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, Oxford, 2002.
- [Jackendoff, 2007] Ray Jackendoff. *Language, Consciousness, Culture: Essays on Mental Structure*. The MIT Press, 2007.
- [Jackendoff, 2009] Ray Jackendoff. Parallels and nonparallels between language and music. *Music Perception*, 26(3):195–204, 2009.
- [Jackendoff, 2011] Ray Jackendoff. What is the human language faculty?: Two views. *Language*, 87(3):586–624, 2011.
- [Jentschke and Koelsch, 2009] Sebastian Jentschke and Stefan Koelsch. Musical training modulates the development of syntax processing in children. *Neuroimage*, 47(2):735–744, 2009.
- [Katz and Pesetsky, 2009] Jonah Katz and David Pesetsky. The identity thesis for language and music. *Draft published online, : lingBuzz/000959*, 2009.
- [Keller *et al.*, 2007] Peter E Keller, Günther Knoblich, and Bruno H Repp. Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Consciousness and Cognition*, 16(1):102–111, 2007.
- [Keller, 2008] Peter E Keller. Joint action in music performance. In F. Morganti, A. Carassa & G. Riva, editors, *Enacting intersubjectivity: A cognitive and social perspective on the study of interactions*, volume 10, page 205–221. Amsterdam: IOS press, 2008.
- [Kempson *et al.*, 2001] Ruth Kempson, W. Meyer-Viol, and Dov Gabbay. *Dynamic Syntax*. Blackwell, Oxford, 2001.
- [Kempson *et al.*, 2009] Ruth Kempson, Eleni Gregoromichelaki, Matthew Purver, Greg J. Mills, Andrew Gargett, and Christine Howes. How mechanistic can accounts of interaction be? In *Proceedings of the 13th SEMDIAL Workshop on the Semantics and Pragmatics of Dialogue (DiaHolmia)*, pages 67–74, Stockholm, Sweden, June 2009. Royal Institute of Technology (KTH).
- [Kempson *et al.*, 2012] Ruth Kempson, Eleni Gregoromichelaki, and Stergios Chatzikyriakidis. Joint utterances in Greek: their implications for linguistic modelling. In *Proceedings of 33rd Annual Linguistics Meeting “Syntactic Theories and the Syntax of Greek”*. Thessaloniki, 26-27 April 2012, 2012.
- [Kivy, 1998] Peter Kivy. *Authenticities: Philosophical reflections on musical performance*. Cornell University Press, 1998.
- [Kivy, 2002] Peter Kivy. *Introduction to a Philosophy of Music*. Clarendon Press, 2002.
- [Koelsch *et al.*, 2005] Stefan Koelsch, Thomas C Gunter, Matthias Wittfoth, and Daniela Sammler. Interaction between syntax processing in language and in music: an ERP study. *Journal of Cognitive Neuroscience*, 17(10):1565–1577, 2005.
- [Krumhansl and Castellano, 1983] Carol L Krumhansl and Mary A Castellano. Dynamic processes in music perception. *Memory & Cognition*, 11(4):325–334, 1983.
- [Leman, 2008] Marc Leman. *Embodied Music: Cognition and Mediation Technology*. MIT Press, 2008.

- [Lerdahl and Jackendoff, 1983] Fred Lerdahl and Ray S Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, 1983.
- [Lerdahl, 1988] Fred Lerdahl. Tonal Pitch Space. *Music Perception*, pages 315–349, 1988.
- [Lerdahl, 1992] Fred Lerdahl. Cognitive constraints on compositional systems. *Contemporary Music Review*, 6(2):97–121, 1992.
- [Lerdahl, 1997] Fred Lerdahl. Composing and listening: A reply to Nattiez. In I. Delige & J. Sloboda, editors, *Perception and cognition of music*, pages 421–428. Psychology Press, 1997.
- [Lerdahl, 2009] Fred Lerdahl. Genesis and architecture of the GTTM project. *Music Perception: An Interdisciplinary Journal*, 26(3):187–194, 2009.
- [Levinson, 1997] Jerrold Levinson. *Music in the Moment*. Cornell University Press, 1997.
- [Levinson, 2000] Stephen C. Levinson. *Presumptive Meanings*. MIT Press, Cambridge, MA, 2000.
- [Levinson, 2012] Stephen C Levinson. Action formation and ascription. In T. Stivers & J. Sidnell, editors, *The Handbook of Conversation Analysis*, pages 101–130. Wiley Online Library, 2012.
- [Levinson, 2013] Stephen C Levinson. Cross-cultural universals and communication structures. In Michael A Arbib, editor, *Language, Music, and the Brain: A Mysterious Relationship*. MIT Press, 2013.
- [Levy, 2008] Roger Levy. Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177, 2008.
- [Linell, 2009] Per Linell. *Rethinking language, mind, and world dialogically: Interactional and contextual theories of human sense-making*. Greenwich, CT: Information Age Publishing, 2009.
- [Lobina, 2011] David J Lobina. A running back and forth: A review of recursion and human language. *Biolinguistics*, 5(1-2):151–169, 2011.
- [Maess *et al.*, 2001] Burkhard Maess, Stefan Koelsch, Thomas C Gunter, and Angela D Friederici. Musical syntax is processed in Broca’s area: an MEG study. *Nature Neuroscience*, 4(5):540–545, 2001.
- [Marin, 2009] Manuela M Marin. Effects of early musical training on musical and linguistic syntactic abilities. *Annals of the New York Academy of Sciences*, 1169(1):187–190, 2009.
- [Marr, 1982] David Marr. *Vision: A computational approach*. Freeman & Co., San Francisco, 1982.
- [McDermott, 1978] Drew McDermott. Planning and acting. *Cognitive Science*, 2(2):71–109, 1978.
- [Meyer, 1956] Leonard B Meyer. *Emotion and Meaning in Music*, pages 256–272, 1956.
- [Meyer, 1973] Leonard B Meyer. *Explaining Music: Essays and Explorations*. University of California Press, 1973.
- [Mills and Gregoromichelaki, 2010] Gregory Mills and Eleni Gregoromichelaki. Establishing coherence in dialogue: sequentiality, intentions and negotiation. In *Proceedings of SemDial (PozDial)*, 2010.
- [Mills, 2011] Gregory J Mills. The emergence of procedural conventions in dialogue. In *Proceedings of the 33rd annual conference of the Cognitive Science Society*, pages 471–476, 2011.
- [Mills, 2013] Gregory J Mills. Dialogue in joint activity: complementarity, convergence and conventionalization. *New Ideas in Psychology*, 2013.
- [Molnar-Szakacs and Overy, 2006] Istvan Molnar-Szakacs and Katie Overy. Music and mirror neurons: from motion to ‘emotion’. *Social Cognitive and Affective Neuroscience*, 1(3):235–241, 2006.
- [Monson, 1996] Ingrid Monson. *Saying Something: Jazz Improvisation and Interaction*. University of Chicago Press, 1996.

- [Narmour, 1977] Eugene Narmour. *Beyond Schenkerism: The need for alternatives in music analysis*. University of Chicago Press Chicago, 1977.
- [Narmour, 1988] Eugene Narmour. On the relationship of analytical theory to performance and interpretation. In E. Narmour & R. A. Solie, editors, *Explorations in music, the arts, and ideas: Essays in honor of Leonard B. Meyer*, pages 317–40, Stuyvesant: Pendragon Press, 1988.
- [Narmour, 1992] Eugene Narmour. *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. University of Chicago Press, 1992.
- [Newmeyer, 2010] Frederick J Newmeyer. What conversational English tells us about the nature of grammar: A critique of Thompson’s analysis of object complements. In K. Boye, & E. Engberg-Pedersen, editors, *Language Usage and Language Structure*, pages 3–44, Berlin: Mouton De Gruyter, 2010.
- [Overy and Molnar-Szakacs, 2009] Katie Overy and Istvan Molnar-Szakacs. Being together in time: musical experience and the mirror neuron system. *Music Perception*, 26(5):489–504, 2009.
- [Pastra and Aloimonos, 2012] Katerina Pastra and Yiannis Aloimonos. The minimalist grammar of action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1585):103–117, 2012.
- [Patel *et al.*, 2008] Aniruddh D Patel, John R Iversen, Marlies Wassenaar, and Peter Hagoort. Musical syntactic processing in agrammatic broca’s aphasia. *Aphasiology*, 22(7-8):776–789, 2008.
- [Patel, 2003] Aniruddh D. Patel. Language, music, syntax and the brain. *Nature Neuroscience*, 6(7):674–681, 2003.
- [Patel, 2008] Aniruddh D. Patel. *Music, Language, and the Brain*. Oxford University Press, Oxford, 2008.
- [Pearce and Wiggins, 2006] Marcus T Pearce and Geraint A Wiggins. Expectation in melody: The influence of context and learning. *Music Perception*, 23(5):377–405, 2006.
- [Pearce and Wiggins, 2012] Marcus T Pearce and Geraint A Wiggins. Auditory expectation: The information dynamics of music perception and cognition. *Topics in Cognitive Science*, 4(4):625–652, 2012.
- [Pearce *et al.*, 2010] Marcus T Pearce, María Herrojo Ruiz, Selina Kapasi, Geraint A Wiggins, and Joydeep Bhattacharya. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage*, 50(1):302–313, 2010.
- [Phillips, 1996] Colin Phillips. *Order and structure*. PhD thesis, Massachusetts Institute of Technology, 1996.
- [Pickering and Garrod, 2004] Martin Pickering and Simon Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–226, 2004.
- [Pickering and Garrod, 2007] Martin Pickering and Simon Garrod. Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3):105–110, 2007.
- [Pickering and Garrod, 2013] Martin J. Pickering and Simon Garrod. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36:329–347, 8 2013.
- [Pinker, 1997] Steven Pinker. *How the Mind Works*. WW Norton & Company, 1997.
- [Piwek, 2011] Paul Piwek. Dialogue structure and logical expressivism. *Synthese*, 183(1):33–58, 2011.
- [Poesio and Rieser, 2010] Massimo Poesio and Hannes Rieser. Completions, coordination, and alignment in dialogue. *Dialogue and Discourse*, 1:1–89, 2010.
- [Pulvermüller and Fadiga, 2010] Friedemann Pulvermüller and Luciano Fadiga. Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5):351–360, 2010.
- [Pulvermüller, 1999] Friedemann Pulvermüller. Words in the brain’s language. *Behavioral and Brain Sciences*, 22:253–279, 1999.

- [Pulvermüller, 2010] Friedemann Pulvermüller. Brain embodiment of syntax and grammar: Discrete combinatorial mechanisms spelt out in neuronal circuits. *Brain and Language*, 112(3):167–179, 2010.
- [Purver *et al.*, 2009] Matthew Purver, Christine Howes, Eleni Gregoromichelaki, and Patrick G. T. Healey. Split utterances in dialogue: A corpus study. In *Proceedings of the 10th Annual SIGDIAL Meeting on Discourse and Dialogue (SIGDIAL 2009 Conference)*, pages 262–271, London, UK, September 2009. Association for Computational Linguistics.
- [Repp and Knoblich, 2004] Bruno H Repp and Günther Knoblich. Perceiving action identity: How Pianists Recognize Their Own Performances. *Psychological Science*, 15(9):604–609, 2004.
- [Rizzolatti and Craighero, 2004] Giacomo Rizzolatti and Laila Craighero. The mirror-neuron system. *Annual Review of Neuroscience*, 27:169–192, 2004.
- [Rizzolatti and Craighero, 2007] Giacomo Rizzolatti and Laila Craighero. Language and mirror neurons. In G. Gaskell, editor, *Oxford Handbook of Psycholinguistics*, Oxford University Press, Oxford, 2007.
- [Rohrmeier *et al.*, to appear] Martin Rohrmeier, Zoltan Dienes, Xiuyan Guo & Qiufang Fu. Implicit learning of recursion. In F. Lowenthal & L. Lefebvre, editors, *Language and Recursion*, Springer Verlag, to appear.
- [Ryle, 1949] Gilbert Ryle. *The Concept of Mind*. University of Chicago Press, 1949.
- [Sammler *et al.*, 2009] D Sammler, S Koelsch, T Ball, A Brandt, CE Elger, AD Friederici, M Grigutsch, H-J Huppertz, TR Knösche, J Wellmer, G Widman, A Schulze-Bonhage. Overlap of musical and linguistic syntax processing: intracranial ERP evidence. *Annals of the New York Academy of Sciences*, 1169(1):494–498, 2009.
- [Sawyer, 2003] R Keith Sawyer. *Group creativity: Music, Theater, Collaboration*. Psychology Press, 2003.
- [Sawyer, 2008] R Keith Sawyer. *Group Genius: The Creative Power of Collaboration*. Basic Books, 2008.
- [Schegloff, 2007] Emanuel A Schegloff. *Sequence Organization in Interaction: A Primer in Conversation Analysis I*. Cambridge University Press, 2007.
- [Schenker, 1935/1979] Heinrich Schenker. *Der Frei Satz*, Vienna: Universal Edition, 1935. Published in English as *Free Composition*, translated and edited by E. Oster. Longman, 1935/1979.
- [Schober and Clark, 1989] Michael F. Schober and Herbert H. Clark. Understanding by addressees and overhearers. *Cognitive Psychology*, 21:211–232, 1989.
- [Sebanz *et al.*, 2006] Natalie Sebanz, Harold Bekkering, and Günther Knoblich. Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2):70–76, 2006.
- [Slevc *et al.*, 2009] L Robert Slevc, Jason C Rosenberg, and Aniruddh D Patel. Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, 16(2):374–381, 2009.
- [Small, 1998] Christopher Small. *Musicking: The meanings of performing and listening*. Wesleyan, 1998.
- [Sperber and Wilson, 1995] Dan Sperber and Deirdre Wilson. *Relevance: Communication and Cognition*. Blackwell, second edition, 1995.
- [Steedman, 1984] Mark J Steedman. A generative grammar for jazz chord sequences. *Music Perception*, pages 52–77, 1984.
- [Steedman, 1996] Mark J Steedman. The blues and the abstract truth: Music and mental models. In A. Garnham & J. Oakhill, editors, *Mental Models in Cognitive Science*, pages 305–318, Erlbaum, Mahwah, NJ, 1996.
- [Steedman, 2000] Mark J Steedman. *The Syntactic Process*. MIT Press, Cambridge, MA, 2000.
- [Steedman, 2002] Mark J Steedman. Plans, affordances, and combinatory grammar. *Linguistics and Philosophy*, 25(5-6):723–753, 2002.

- [Steinbeis and Koelsch, 2008] Nikolaus Steinbeis and Stefan Koelsch. Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, 18(5):1169–1178, 2008.
- [Suchman, 2007] Lucille Alice Suchman. *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press, 2007.
- [Sun *et al.*, 2005] Ron Sun, L Andrew Coward, and Michael J Zenzen. On levels of cognitive modeling. *Philosophical Psychology*, 18(5):613–637, 2005.
- [Tettamanti and Moro, 2012] Marco Tettamanti and Andrea Moro. Can syntax appear in a mirror (system)? *Cortex*, 48(7):923–935, 2012.
- [Thompson-Schill *et al.*, 2013] Sharon Thompson-Schill, Peter Hagoort, Peter Ford Dominey, Henkjan Honing, Stefan Koelsch, D. Robert Ladd, Fred Lerdahl, Stephen C. Levinson S, and Mark Steedman. Multiple levels of structure in language and music. In Michael A Arbib, editor, *Language, Music, and the Brain: A Mysterious Relationship*. MIT Press, 2013.
- [Tsoulas, 2010] George Tsoulas. Computations and interfaces: some notes on the relation between the language and the music faculties. *Musicae scientiae*, Discussion Forum 5:11–41, 2010.
- [Widmer, 1995] Gerhard Widmer. Modeling the rational basis of musical expression. *Computer Music Journal*, 19(2):76–96, 1995.
- [Wiggins *et al.*, 2010] Geraint Wiggins, Daniel Müllensiefen, Marcus T Pearce. On the non-existence of music: Why music theory is a figment of the imagination. In *Musicae Scientiae, Discussion Forum*, volume 5, pages 231–255. ESCOM, 2010.
- [Wimsatt and Beardsley, 1946] William K Wimsatt and Monroe C Beardsley. The Intentional fallacy. *The Sewanee Review*, 54(3):468–488, 1946.
- [Wolpert *et al.*, 2003] Daniel M Wolpert, Kenji Doya, and Mitsuo Kawato. A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431):593–602, 2003.

Eleni Gregoromichelaki  
King’s College London, UK.  
eleni.gregor@kcl.ac.uk