Received: 8 May 2018 | Revised: 20 September 2018 | Accepted: 2 October 2018

DOI: 10.1002/wcs.1484

WILEY WIREs COGNITIVE SCIENCE

ADVANCED REVIEW

Shared neural and cognitive mechanisms in action and language: The multiscale information transfer framework

Alice Blumenthal-Dramé[1,3] | Evie Malaia[2,3]

[1]Department of English, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany

[2]Department of Communicative Disorders, University of Alabama, Tuscaloosa, Alabama

[3]Freiburg Institute for Advanced Studies, Freiburg, Germany

Correspondence
Alice Blumenthal-Dramé, Department of English, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany.
Email: alice.blumenthal@anglistik.uni-freiburg.de

Funding information
European Union Marie S. Curie FRIAS COFUND Fellowship Programme; U.S. National Science Foundation, Grant/Award Number: 1734938

This review compares how humans process action and language sequences produced by other humans. On the one hand, we identify commonalities between action and language processing in terms of cognitive mechanisms (e.g., perceptual segmentation, predictive processing, integration across multiple temporal scales), neural resources (e.g., the left inferior frontal cortex), and processing algorithms (e.g., comprehension based on changes in signal entropy). On the other hand, drawing on sign language with its particularly strong motor component, we also highlight what differentiates (both oral and signed) linguistic communication from nonlinguistic action sequences. We propose the multiscale information transfer framework (MSIT) as a way of integrating these insights and highlight directions into which future empirical research inspired by the MSIT framework might fruitfully evolve.

This article is categorized under:
Psychology > Language
Linguistics > Language in Mind and Brain
Psychology > Motor Skill and Performance
Psychology > Prediction

KEYWORDS
action, communication, information theory, language, signal processing

1 | INTRODUCTION

The ability to perceive and interpret signals from the environment, and to construct mental representations using those signals, is a defining feature of human communication. Outside the domain of intentional communication, the ability to parse continuous information into events, and to incorporate these events into schemas (i.e., memory templates constructed on the basis of prior experience) is what underscores all perception, understanding, and action planning (Zacks & Tversky, 2001). Despite striking similarities between research on information extraction in action observation and language comprehension, there has been little research connecting the two. The bodies of literature on the analysis of signal parameters in both domains, as well as their neural processing, have very little overlap.

One way of uniting these domains of inquiry is through the multiscale information transfer framework (MSIT), which incorporates a continuum from the potentially informative signal (e.g., human motion, sign language communication, or the auditory/visual signal in a known or unknown language) to its neural processing, with the goal of quantifying and predicting how potentially interpretable information in the signal (quantitatively defined) is processed and converted into mental representations, linguistic or nonlinguistic. The MSIT framework is the first attempt to unite cognitive neuroscience and signal processing research on language comprehension and action observation. So far, these strands of research have largely evolved in parallel without much interaction and cross-fertilization in spite of striking similarities in terms of interpretation and modeling

 © 2018 Wiley Periodicals, Inc.

of overlapping cognitive processes. This paper aims to account for the following questions: First, which cognitive mechanisms (e.g., perceptual segmentation, predictive processing, integration across multiple temporal scales), neural resources (e.g., the left inferior frontal cortex), and processing algorithms (e.g., comprehension based on changes in signal entropy) do humans rely on to understand information from their environment? Second, how does the rate of information transfer allowed by (oral and signed) linguistic sequences differ from that allowed by nonlinguistic action sequences?

Our review of the literature is organized to describe the separate components of the framework which is presented in Figure 1.

The MSIT is a modality-independent framework bringing together research on action understanding and event segmentation on the one hand, and signed and spoken language processing on the other hand. The MSIT accounts for the hierarchical structuring of information across multiple scales, and allows for the selection of mathematical modeling tools to analyze information extraction at specific levels of spatiotemporal resolution.

We start by showing that both action and language sequences can be interpreted as exhibiting an analogous, hierarchically scaled structure. We go on to review the literature on the segmentation of naturalistic dynamic scenes and of (signed and oral) language, with the goal of relating the processes underlying signal parsing and meaning extraction to current understanding of neural engagement in both processes. We then identify promising approaches to the computational modeling of a continuum that includes the communicative signal and its neural processing, taking into account the multiscale nature of the communicative signal and the system processing it—the human brain. Finally, we review recent research suggesting that the understanding of both action and language sequences draws heavily on predictive processing operating across the multiscale hierarchy.

2 | HIERARCHICAL STRUCTURE IN LANGUAGE AND ACTION

A key tenet of embodied cognition is that human higher-order cognition (notably planning, reasoning, and language) is grounded in emotional, affective, perceptual and motor experiences with the world. Within the embodied cognition framework, theories of motor cognition have focused on the more specific claim that higher-order cognition exploits cognitive mechanisms and neural circuits also used for the execution of goal-directed motor sequences. This has led to the suggestion that language should be considered as an internalized form of goal-directed action whose execution is suppressed (Barsalou, 2008; Goldinger, Papesh, Barnhart, Hansen, & Hout, 2016; Grafton, 2009). Based on research into the relationship between action and language, the view that both phonetic and semantic processing shares cognitive and neural resources with the sensorimotor system has gained increasing acceptance (Barsalou, 2016; Fischer & Zwaan, 2008; Willems & Hagoort, 2007). For example, comprehending the meanings of words semantically related to body parts (e.g., *eat*, *mouth*) activates brain regions also engaged in moving these body parts(Andrews, Frank, & Vigliocco, 2014; Meteyard, Cuadrado, Bahrami, & Vigliocco,

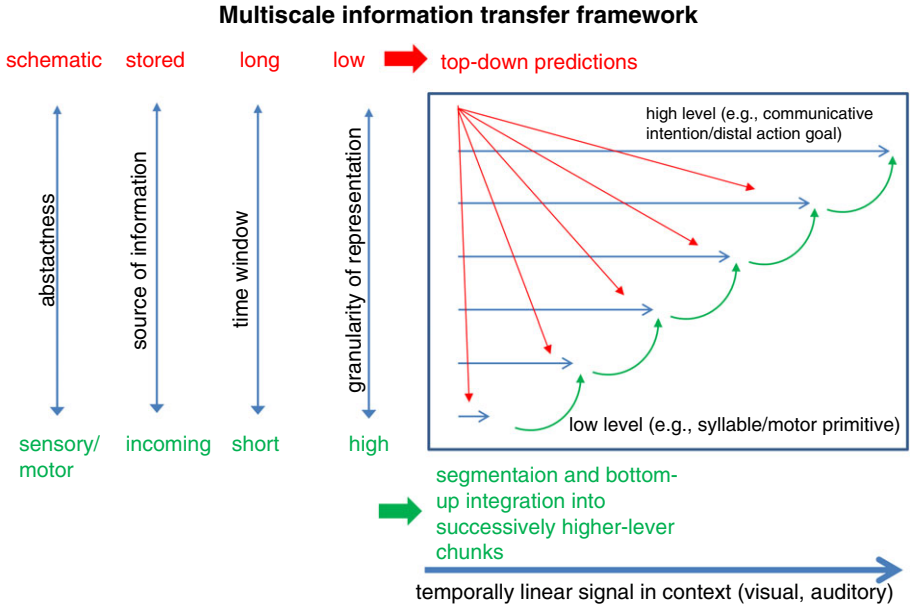


FIGURE 1 Model detailing the multiscale information transfer framework. The sensory signal, which unfolds linearly in time, contains parameters that are processed at multiple scales of resolution in both language comprehension and action observation. The incoming sensory signal is segmented into chunks at multiple scales under the top-down guidance of the processor's predictions

2012; Pulvermüller, 2013), just like decoding sounds recruits regions responsible for the production of these sounds (Nuttall, Kennedy-Higgins, Devlin, & Adank, 2017; Skipper, Devlin, & Lametti, 2017).

However, a question that has only recently started to receive scholarly attention is whether the boundaries of sensorimotor involvement in language processing can be extended beyond phonetics and semantics to encompass all hierarchical levels of language. More specifically, does the reliance on motor circuits also generalize to the processing of language syntax? A positive answer to this question would be at odds with established opinion, which has treated syntax as a qualitatively distinctive module in the human brain (Everaert, Huybregts, Chomsky, Berwick, & Bolhuis, 2015; Moro, 2014a, 2014b; Tettamanti & Moro, 2012), rather than a skill lying on a continuum with nonlinguistic human and animal skills. In support of this traditional view, it has been argued that the abstract hierarchical structure of syntax is, by definition, not directly accessible to the sensorimotor system. Moreover, motor sequences have been pointed out to lack essential features of language, such as communicative function, a lexicon including open and closed-class items, as well as structures involving phrase-structure hierarchies and recursion.

By contrast, proponents of the embodied framework have described language (including syntax) as a type of sensorimotor skill (Chater, McCauley, & Christiansen, 2016; Pickering & Garrod, 2013; Pulvermüller, 2014; Pulvermüller & Fadiga, 2010). The present section reviews recent literature showing that the conceptual gap between action and language syntax is easier to bridge than traditionally assumed, with action and language sequences being structured in analogous, hierarchical terms.

Thus, while both language and goal-directed actions are realized as linear sequences, they are driven by plans exhibiting hierarchical structure (Boeckx & Fujita, 2014; Clark, 2013; MacDonald, 2013). In this context, the notion of hierarchy refers to a multilayered representation where each higher-order level incorporates several units from the immediately preceding level. In both language and motor sequences, higher levels operate over a larger time windows, since they encode more abstract and distal goals. By contrast, lower levels represent more fine-grained, proximal, and concrete execution/perception details (Diedrichsen & Kornysheva, 2015; Garrod, Gambi, & Pickering, 2014; Grafton & Hamilton, 2007). It must be emphasized from the outset that this notion of hierarchy is not related to the formal linguistic notion of phrase-structure hierarchy (Fitch & Martins, 2014; Martins, Martins, & Fitch, 2016).

2.1 | Constituency levels in motor and language hierarchies

The linguistic hierarchy is comprised of units of different levels of granularity from sounds up to discourse structures (e.g., phone, syllables, morphemes, lexemes, phrases, sentences) (Christiansen & Chater, 2016; Clark, 2013; Garrod et al., 2014; Kuperberg & Jaeger, 2016). The motor hierarchy encodes the velocity and trajectory of minimal action components at its lowest level (often referred to as the "kinematic level"), and distal intentions at its highest level (e.g., to drink milk). Intermediate-level representations include the "motor level," which encodes patterns of muscle activity, and the immediately superior "goal level," which specifies sub-goals leading up the realization of the overall intention (e.g., grasp a glass, open the fridge, grasp a milk carton, open it, etc., are sub-goals with regard to the overarching intention to drink milk) (Diedrichsen & Kornysheva, 2015; Kilner, 2011). Each sub-goal can itself be subdivided into even more proximal sub-constituents (e.g., extend the arm, form a power grip, grasp the glass from above or from the side, etc., to grasp a glass).

Hierarchical structure has been interpreted as a domain-general cognitive response to memory constraints and interference effects between low-level units, which make it impossible to plan and keep online complete low-level sequences (Christiansen & Chater, 2016; MacDonald, 2013). Hierarchical representation allows sequence producers to convert global high-level plans into successively more detailed and specific low-level representations, with the most concrete representations being generated immediately before production. Hierarchical structure does not only allow producers to deal with memory and interference constraints, but also to recombine lower-level items in a flexible manner while maintaining fluency (Christiansen & Chater, 2016). As a result, individual steps can be modified or omitted, which would not be possible if sequences were driven by purely sequential association links (Diedrichsen & Kornysheva, 2015; MacDonald, 2013; Martins et al., 2016). For example, in getting dressed, clothes may be put on in different (although by no means fully random) orders. Likewise, certain sub-chunks of language sequences can be moved around. At the same time, both kinds of sequences are subject to similar constraints (e.g., agents need to extend their arm before grasping a cup, and determiners must precede the noun in English noun phrases, cf. *the dog* vs. **dog the*) and show similar effects of practice and repetition (cf. Box 1).

Hierarchically organized language and motor sequences draw on common neural circuits. Within these circuits, which are generally attributed to the left perisylvian cortex, Broca's area (i.e., broadly premotor area BA 44 and BA 45) has been singled out as a prime candidate for containing a processor critically engaged in the production and comprehension of hierarchically structured sequences across different modalities, including language and motor sequences (Clerget, Andres, & Olivier, 2013; Clerget, Winderickx, Fadiga, & Olivier, 2009; Fadiga, Craighero, & D'Ausilio, 2009; Fitch & Martins, 2014; Jeon, 2014; Koechlin & Jubault, 2006; Pulvermüller & Fadiga, 2010; Tettamanti & Weniger, 2006). Thus, syntactic comprehension deficits

BOX 1

EFFECTS OF REPETITION ACROSS DOMAINS

The processing of nonautomatized sequences in both modalities involves the selection of individual low-level units (or "primitives"), and their combination into a linear chain (in this context, linguistic "primitives" refer to words or morphemes, whereas motor "primitives" refer to "spatiotemporal patterns of coordinated muscle activity" that are stable across different complex movements, cf. Diedrichsen & Kornysheva, 2015, p. 227) (Pastra & Aloimonos, 2012). Just think of learning how to tie one's shoe laces, which initially involves the conscious selection of each individual sub-step of the sequence and their effortful concatenation.

In both types of sequences, automatization through repetition results in the generation of chunks. Chunk status results from a tightening of sequential bonds between the sub-components of a sequence. It implies the selection of the whole chain as a single prefabricated unit, which allows for enhanced ease, fluency and accuracy in online processing (Blumenthal-Dramé, 2012, 2016a; Diedrichsen & Kornysheva, 2015). Moreover, chunk status results in "emancipation." This means that the mental representation for the primitives making up a chunk becomes independent from the representation of the same items in other (i.e., nonchunked) contexts. In language, this loss of component identity leads to well-documented effects of phonological reduction (cf. *going to* > *gonna*). Across language and motor sequences, evidence for chunk status includes a higher probability of errors and larger time lags at the onset of chunks and between chunks (compared to within chunks), reduced variability in the execution of the sequence, fusion of the elements of the sequence, and a reduced duration and intensity of muscular activity (Arnold, Wing, & Rotshtein, 2017; Blumenthal-Dramé et al., 2017; Bybee, 2003; Thompson, McColeman, Stepanova, & Blair, 2017). Chunk status does not only influence the production of sequences in different modalities; it also modulates how observers and comprehenders segment the incoming sensory signal. For example, high-frequency chunks are both more difficult to segment and more expected than lower-frequency sequences (Blumenthal-Dramé, 2012, 2016a, 2016b; Blumenthal-Dramé et al., 2017).

When human accumulate experience with sequences that are perceived as similar, but not identical, this leads to the generation of a "schema." A schema can be thought of as a memory unit which generalizes across similarities, but abstracts away from differences between sequences of analogous structure. Schemas allow for some degree of flexibility regarding the concrete execution of sequences, thereby allowing for transfer and adaptation to novel contexts (Braun, Mehring, & Wolpert, 2010; Goldberg, 2016). Thus, experience in inline skating will facilitate the learning of ice-skating, and experience with different sentences conveying a resultative meaning (e.g., *He painted his house pink*) will facilitate the generation of previously unheard, but structurally and functionally analogous sentences (e.g., *Sally sneezed the napkin off the table*; Goldberg, 1995, p. 6). Specific parameters of the incoming signal can trigger the extraction of an event schema from long-term memory, and its application to interpret incoming data (Kurby & Zacks, 2008; Malaia, Wilbur, & Weber-Fox, 2013). This process is remarkably flexible: schema transfer has been attested between action and language (Klima et al., 1999; Strickland et al., 2015).

induced by brain lesions in the left perisylvian cortex have been shown to correlate with impairments in the high-level sequencing of certain complex actions and utterances (Fazio et al., 2009). It has been suggested that within the left inferior frontal area, anterior parts (notably BA47/45) may participate in networks underlying the encoding of high hierarchical levels, whereas more posterior parts (BA 44/BA6) subserve the encoding of lower-level goals (Clerget et al., 2013; Kilner, 2011).

3 | MULTISCALE INPUT PARSING

How do we, as humans, make sense of our surroundings? Although reality is continuous, humans perceive and understand it as consisting of discrete events (Zacks, 2004; Zacks, Speer, Swallow, Braver, & Reynolds, 2007; Zacks & Tversky, 2001; Zacks, Tversky, & Iyer, 2001). To illustrate this on a simple example, a morning family breakfast can go something like this for a parent: cook the breakfast, serve it, invite everyone to the table, deal with the quarrel of children over who gets to use the jar of favorite jam first, eat, make more coffee, dismiss the children, relax for a few minutes, clean up. Humans base their understanding and interpretation of what happens around them on perceptual information that comes in the form of visual, auditory, and other signals, as well as the hierarchical scenarios that they can imagine. The two streams of information—the external signal and the internal predictions for the possible continuation of the signal—interact in the creation of the current model of reality, on which to base behavior. Action and language are similar in relying on models at different levels of scaling, as discussed in the previous section. In this section, we will focus our attention on the similarities between action and sign

language parsing that arise from domain-specific features (e.g., reliance on the visual signal), as well as the differences between action and language parsing.

3.1 | Perceptual segmentation of the visual signal

In the realm of event segmentation, multiple studies have shown that reality is segmented into events at multiple scales simultaneously (Zacks, 2004; Zacks et al., 2001; Zacks, Tversky, et al., 2001). Event segmentation studies typically ask that participants watch a video with a dynamic scene and indicate (e.g., via button-press) the time-points at which they think an action is completed; the participants do so at either fine-grained or coarse-grained boundaries. The temporal relationship between coarse- and fine-grained boundaries reveals the hierarchical structure of fine-grained events within coarse-grained ones (Zacks, Tversky, et al., 2001). A recurring finding is that participants are in remarkable agreement with regard to the boundaries of coarse and fine events, both in realistic scenarios (which could be based on top-down understanding of how one folds laundry), and in abstract moving-object experiments (Kurby & Zacks, 2008; Speer, Zacks, & Reynolds, 2007; Zacks, Braver, et al., 2001). In moving-object studies, participants often attribute intentions and goals to moving objects as justification for segmentation of events at longer temporal scales. As humans do not attend to all of the information that is available through perceptual channels, the signal features that participants attend to in moving-object experiments are of particular interest for understanding how event segmentation proceeds. From multiple features which were experimentally tested as potentially relevant for event segmentation (e.g., distance between pairs of moving objects, relative location, speed, acceleration, etc.), changes in the speed of individual objects emerged as the feature most highly correlated with event boundary identification. Specifically, event onsets and offsets were timed with increase and decrease of speed (or acceleration and deceleration parameters). At the neural level, these changes were associated with increased activity in the visual motion area MT+ area, as well as a nearby region in the superior temporal sulcus associated with the processing of biological motion (Zacks, Swallow, Vettel, & McAvoy, 2006). Very similar neural activations are observed in sign-naïve participants observing sign language sentences (Malaia, Ranaweera, Wilbur, & Talavage, 2012); yet, signers observing the same stimuli show only language-processing activation. This indicates that while all participants operate on the same perceptual information, it is only familiarity with the language which allows low-level perception of motion differences in the signal to be transferred as information to higher, language-based processing scales (e.g., phonology, semantics, and syntax).

The results of research on visual (sign) language and action understanding can be interpreted within the MSIT framework in the following way: In the absence of top-down scenarios, it is the relative predictability of the signal that appears to mark the boundaries between events. When an object picks up speed, it is difficult to say where in space the object might end up being located at a specific point in time. When the speed of the object drops, the location of the object in time is easier to predict; the variability of predicted outcomes falls, and an event boundary is identified. However, signers process the predictability of the signal in a different manner. For them, the phonotactic constraints of the sign language, the lexical constraints of the lexicon, and the syntactic constraints all operate on the signal at the same time, such that motion differences are translated into linguistic features in the visual domain (Krebs, Malaia, Wilbur, & Roehm, 2018). Thus, the relationship between the signal and its interpretation is more complicated in sign language perception, as the interpretation takes place at more levels (lexical, syntactic, prosodic), as compared to the perception of motion (Malaia, Borneman, & Wilbur, 2008; Malaia & Wilbur, 2012b; Malaia, Wilbur, & Weber-Fox, 2013). This is not to say that motion in actions is interpreted at only one level of meaning. For example, if a baby throws a rattle, a parent can identify not just the event of rattle-throwing, but also the emotion (angry, playful, accidental) underlying the action. In sign language motion, emotional content can also be overlaid with information content; but the information content of motion would also be interpretable as containing several levels of meaning, such as semantic, syntactic, and prosodic information.

3.2 | Multiscale segmentation in sign language

At conceptual linguistic levels, event segmentation is encoded by the feature of semantic telicity. Telicity is a feature of the verb or verb phrase which identifies whether the event described has an identifiable boundary. In English, verbs like “drop” or “accomplish” are telic; verbs that describe actions (“read”) or states (“sleep”) are atelic. In sign languages, the semantics of verb phrases related to the telicity structure of the described event is traceable to the motion of hand articulators. Motion capture studies (Malaia & Wilbur, 2012a; Malaia, Wilbur, & Milkovic, 2013) of two unrelated sign languages—American Sign Language (ASL) and Croatian Sign Language (HZJ) have indicated that in both languages, velocity and acceleration of the dominant hand motion during the production of the verb sign was related to linguistic telicity, that is, allowed to identify events as segmentable or ongoing. However, articulator motion in both languages was also subject to influence by other components of the linguistic system.

In motion capture investigations of ASL and HZJ, participants were asked to produce sentences in which the verbs of interest (targets) were located either in the middle of sentences, or at their end (either placement is allowed by the syntactic rules of both languages). Analysis of 3D displacement of the signers' wrists during verb articulation indicated that motion was slower when verbs were signed at the end of sentences. This phenomenon is known in linguistic research as Phrase Final Lengthening—a prosodic process that has been identified across signed and spoken languages. Thus, the physical features of the motion signal were not only influenced by verb semantics, but also by sentential prosody.

In addition to this, verbal telicity marked by articulator deceleration was shown to be differently situated within the linguistic system of each sign language (Malaia, Wilbur, & Milkovic, 2013), such that different levels of the linguistic system interact in processing telicity, even though it is marked in the same way: by visual velocity of dominant hand motion. In ASL, motion deceleration in the verb to mark telicity or to communicate an event boundary is a semantic feature: It is usually not possible to modify motion in atelic verbs to impart a meaning of event boundary to the verb phrase. By contrast, in HZJ, the acceleration of the dominant (and sometimes nondominant) hand is part of the morphological system: Most verb signs can be modified to indicate an event boundary.[1]

To summarize, in the case of sign languages, differences in the relative speed of motion in the visual signal are processed at multiple levels: phonological (as syllable structure differences) (Malaia, Ranaweera, et al., 2012), semantic (as event structure), syntactic (as affecting the patient of the event or not) (Malaia, Wilbur, & Di Sciullo, 2012), and prosodic (as the motion of the verb is affected by its location in the sentence and is subject to phrase-final lengthening) (Malaia & Wilbur, 2012b).

These differences in the relationships between motion parameters in signed events and various components of the linguistic system in different sign languages highlight one important feature of signal processing for comprehension: the temporally linear signal carries information that is segmentable simultaneously at different scales, and is meaningful for different components of the linguistic system. As ASL is a predominantly monosyllabic language (Wilbur, 2009), it most directly demonstrates the similarities between visual event segmentation and linguistic event segmentation, which concurrently engages multiple levels of linguistic processing: phonological, semantic, and syntactic. Yet, although the visual nature of the signal unites action and sign language information (Malaia & Wilbur, 2018), the multilevel nature of linguistic processing is equivalent between speech and sign language.

3.3 | Multiscale segmentation in spoken language

In spoken languages, telicity—as a proxy for event segmentation above the perceptual level—also manifests at, and interacts with, multiple levels of the linguistic hierarchy. Thus, event structure can be marked at phonological, semantic, morphological, and syntactic levels. For example, in Japanese, telic verbs are marked by nonlow vowels (Fujimori & Di Sciullo, 2012). Slavic languages use morphology to convey the temporal structure of event segmentation, such that aspect and telicity in Slavic languages often appear fused, with verbs denoting telic events also marked as lexically perfective, and verbs denoting atelic ones—imperfective (Malaia, 2004; Milkovic & Malaia, 2010). Indonesian utilizes a similar morphological method of telicity marking, employing constructions with the suffix *-kan* (Son & Cole, 2008). Telicity can also be construed at higher levels: not just the lexical verb, but the entire verb phrase can be responsible for denoting event structure. In English and other Germanic languages, the sentence "I ate the fish" denotes a single, telic event, while "I ate fish" (without argument quantification of the direct object) does not. Another grammatical means to denote telicity is the differentiated use of auxiliary verbs in suppletive forms (e.g., in German, the use of *haben* (have) vs. *sein* (be) can differentiate telic from atelic events in complex tense forms). However, even when event structure marking can be identified as belonging to a specific level of the language system, online processing and neuroimaging studies clearly indicate that event structure affects multiple processing systems simultaneously. Online processing studies in English and German have indicated that event structure appears to interact with processing costs for both semantic and syntactic parameters of the sentence (Malaia & Newman, 2015a, 2015b; Malaia, Wilbur, & Weber-Fox, 2009, 2012, 2013; Philipp, Graf, Kretzschmar, & Primus, 2017). For example, in German, event structure was shown to interact with the animacy of sentence subjects (telicity was marked by a combination of adverbial, case, and auxiliary verb, respectively, e.g., *hat [has] über [above] dem [the:DAT] Fluss [river] geschwebt [floated]* "has floated above the river" vs. *ist [is] auf [on] den [the:AKK] Acker [field] geschwebt [floated]* "has floated to the field") (Philipp et al., 2017).

The pattern of cross-scale integration between different components of the linguistic system, as described by the MSIT framework, is also observable in neuroimaging research of spoken and signed languages. It is well-documented that when participants are given a purely semantic task, yet presented with stimuli that differ in event structure, syntactic processing takes place (Malaia et al., 2009; Malaia & Newman, 2015a, 2015b; Malaia, Wilbur, & Weber-Fox, 2012). For example, a comparison of functional magnetic resonance imaging (fMRI) data on the processing of telic and atelic verbs in Italian with a verb-matching task showed that participants had increased activity in left posterior middle temporal gyrus, the brain area typically involved in the processing of argument structure (Romagno, Rota, Ricciardi, & Pietrini, 2012). The authors suggest that event

structure differences between the two classes of stimuli triggered processing of conceptual information relevant to morphosyntax. In other words, conceptual events were processed at multiple scales of linguistic analysis (here, semantics and morphosyntax) simultaneously.

A similar question was investigated by a neuroimaging study of English sentences with telic and atelic verbs in relative clauses (Malaia, Wilbur, & Weber-Fox, 2013). The study orthogonally manipulated verbal telicity in relative clauses and animacy of the first argument.[2] The combination of inanimate subject and reduced relative clause yielded a garden-path effect. Differences in neural activation due to the interaction of telicity and animacy during sentence processing were observed in BA 47, an area also known to support syntactic computation and working memory (Ranganath, Johnson, & D'Esposito, 2003), and the posterior cingulate/precuneus (PCC). The role of PCC in episodic memory retrieval has been supported by a variety of analyses, including lesion studies (Rudge & Warrington, 1991; Valenstein et al., 1987), neuroimaging research (Maddock, Garrett, & Buonocore, 2001; Shannon & Buckner, 2004; Wagner, Shannon, Kahn, & Buckner, 2005), and meta-analyses (Leech & Sharp, 2014; Malaia & Wilbur, 2018; Nielsen, Balslev, & Hansen, 2005). Activation of PCC has also been observed at event boundaries in both neuroimaging studies on perceptual event segmentation (Speer et al., 2007; Speer, Swallow, & Zacks, 2003; Zacks, Braver, et al., 2001), and neurolinguistic investigations of event structure (Malaia & Newman, 2015a, 2015b; Malaia, Ranaweera, et al., 2012; Malaia, Wilbur, & Weber-Fox, 2013). All of these studies searched for the neural mechanisms underlying the cognitive process of segmentation (based either on visual or linguistic input). The overlap in neural activations observed across a battery of studies ties in with the hypothesis that the same neural algorithms underlie segmentation across modalities.[3] For linguistic studies in particular, the neuroimaging evidence combining engagement of BA 47 and PCC suggests that telic verbs facilitate sentence comprehension via the activation of event schemas in episodic memory, and by priming syntactic structures for bounded events. This is compatible with the view that linguistic comprehension relies on the integration of syntactic and event knowledge in a distributed network including language-processing as well as working and episodic (long term) memory regions. This research fits well with the proposed MSIT framework, as it underscores multiscale processing in both language and action processing.

A neuroimaging study of ASL processing provided a conceptual replication of these results, highlighting the similarities of visual action processing in signers and nonsigners, as well as differences due to linguistic processing of the signal by signers only. All participants in the study were presented with video clips of ASL verbs which differed in both the velocity of motion and the event structure of the stimuli: Telic verbs had a quantifiably different deceleration pattern at the end of the sign (Malaia et al., 2008; Malaia, Ranaweera, et al., 2012). In hearing nonsigners, the differences in velocity between verb types elicited increased bilateral activation in the fusiform gyrus, as well as right-hemisphere activation in superior temporal gyrus (STG) and superior parietal lobe. These results suggest that while nonsigners were able to perceive differences in motion features, those differences were too subtle to elicit event schema processing in long-term memory; instead, they were processed purely as spatial differences in motion. ASL signers, on the other hand, processed those differences as purely linguistic: specifically, as related to the syllabic structure of the verbs. As ASL is a predominantly monosyllabic language, differences in velocity of articulator motion are equivalent to differences in the coda of a spoken language syllable (cf. *map—man*). In spoken languages, syllable timing is critically dependent on cerebellar activation (Elliott & Theunissen, 2009; Llanos, Alexander, Stilp, & Kluender, 2017; Stilp & Kluender, 2010). In an ASL neuroimaging study, telic signs elicited activation of a network including the cerebellum, as well as right STG and precuneus (PCC), which, as discussed earlier, are involved in the processing of event schemas. Activation of right STG in signers and nonsigners is also likely indicative of processing differences, as the right hemisphere in signers is routinely involved in the processing of spatial parameters of sign language, to the extent that even resting-state networks in signers show strong right lateralization (Malaia, Talavage, & Wilbur, 2014). Since a significant amount of linguistic processing for sign languages occurs in the right hemisphere, right-hemispheric lateralization is not surprising: Right STG activation in sign language processing is typically associated with the creation of abstract phonological representations based on spatiotemporal properties of the signal.

The findings discussed in this section converge on the understanding that both the parsing of linguistic and nonlinguistic signals and the comprehension of action and language engage a common set of neural networks and rely on similar physical features in the signal. The differences between action and language processing lie in the granularity of scale-dependent processing. Yet, both action and language signals give rise to multiscale processing. The MSIT framework captures this insight from perceptual, processing, and neuroimaging studies, and describes plausible interactions between the external signal, and the multilevel processing of the signal that gives rise to event segmentation or language comprehension.

4 | INFORMATION TRANSFER MEASURES IN COMMUNICATION

Recent methodological and conceptual advances have led to a fundamental reappraisal of how to model the interaction between parameters of the incoming signal, and both the perceptual and cognitive processing of it. Measuring the potential

information contained in the signal using the tools of information theory was one such breakthrough. The mathematical construct of information is easily conceptualized in terms of entropy or variability of the signal over time. For example, if we think of a series of numbers such as [1 0 1 1 0 1 0 1 …], and ask what the likelihood of the next number is, there are only two choices: 1 and 0. The amount of information in the next number in these series is fairly small, since the variability is small. If we use a wider range of numbers, for example, [5 3 0 7 3 2 5 7 8 2 …], the probability of encountering a specific one next drops to 10%. The amount of information transferred per unit of time is, then, defined by the potential variability in the signal over time. In other words, information in the signal is inversely related to the predictability of the signal. Consider a series such as [101 1 11 01 110100 101 …]. Here, the information is conveyed by the numbers, as well as their groupings (groupings constitute a new, hierarchically higher scale, which can be assessed for entropy separately from the raw number sequence).

How can the concept of entropy as an information transfer metric applicable within and across scales be applied to language and speech? In the domain of spoken language, several lines of research converge in identifying temporal changes in the entropy of the speech signal as the basis for information extraction at multiple levels, from linguistic intelligibility to information about accent or speaker gender (Elliott & Theunissen, 2009; Llanos et al., 2017; Stilp, Kiefte, Alexander, & Kluender, 2010; Stilp & Kluender, 2010, 2016; Stilp, Rogers, & Kluender, 2010).

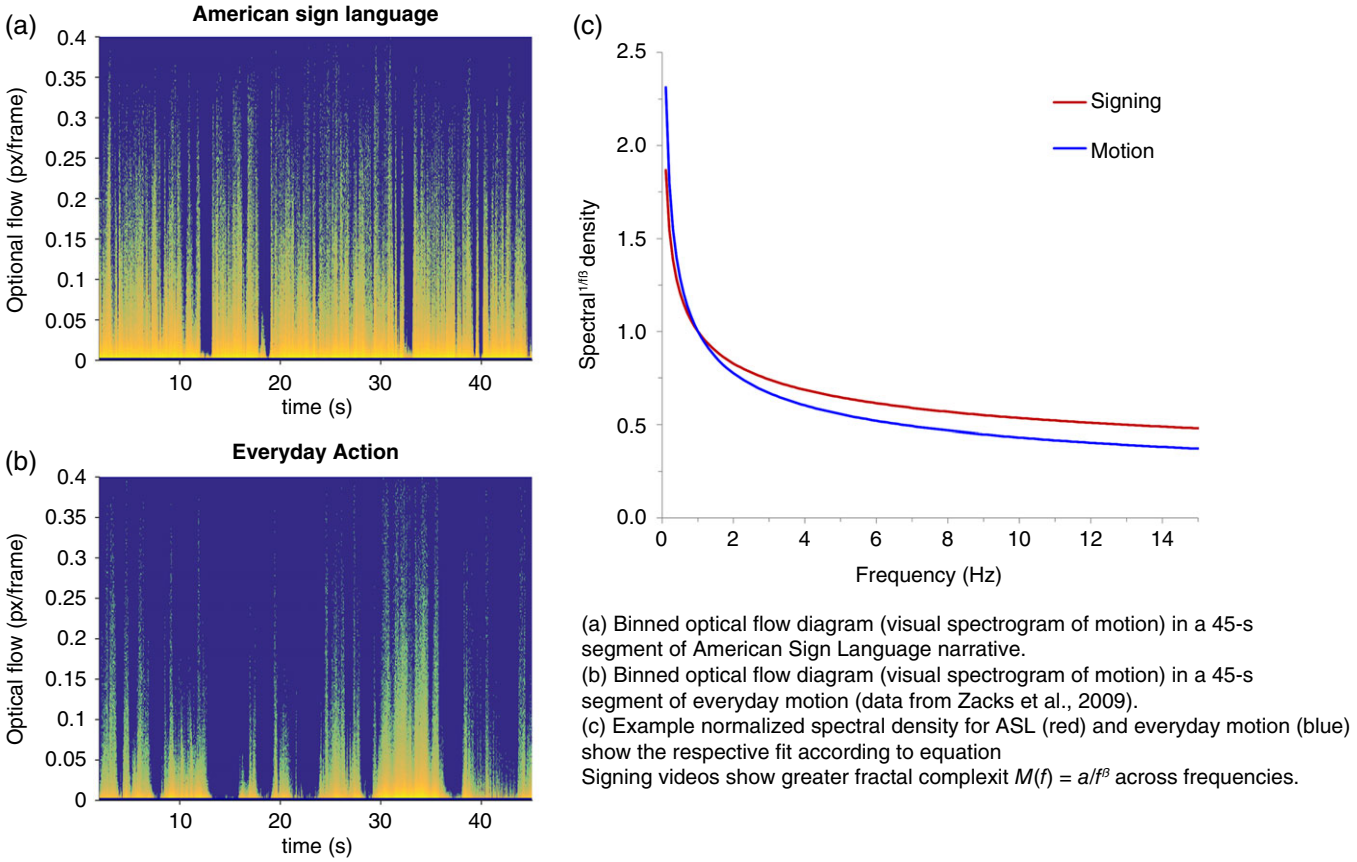
Perceptually, the ability to identify, hierarchically structure, and remember entropy-rich portions of the signal appears to be transferable between action and linguistic domains. The phenomenon of competence transfer from language to action perception has been demonstrated by experiments that compared signers' and nonsigners' perception of dynamic point-light motion (Klima et al., 1999). In a series of studies, participants were asked to view dynamic point-light tracing of pseudo-hieroglyphics, and to draw them as closely to the stimuli as possible. The signers outperformed the nonsigners in the reproduction of the stimuli. In particular, the signers made a crucial distinction between strokes and transitions in the point-light display. Subtle differences between strokes and transitions, or, in other words, changes in signal entropy, were apparent to the signers. Thus, although the stimuli were not linguistically informative for any of the participants; the signing participants were able to extrapolate their linguistic experience with the perceptual segmentation of a signal rapidly varying in visual entropy to a nonlinguistic task that focused on action segmentation and structuring.

The ability to process signal changes is also transferable from action observation to language. For example, sign-language naïve participants who view videos of telic and atelic sign language verbs (which differ in motion signatures), perform remarkably well in correlating event structure between visual and linguistic domains (Strickland et al., 2015).

The observations of perceptual competence transferability between action observation and language fit well with the understanding of multiscale entropy-based processing of input, as formulated within MSIT framework. While event segmentation and sign language parsing appear to use the motion deceleration/acceleration features for parsing the signal, these features are, in essence, markers of entropy in the visual signal. For those with shared linguistic background (sign or speech lexicon and syntax), the syllables are perceptual events that feed into the neurolinguistic processing chain that results in comprehension, whereas for those without appropriate experience, the process ends at the detection of change differences. Understanding the neural algorithms that underlie multiscale information extraction, and developing approaches to measuring comprehension as the use of potential information to construct mental representations, linguistic or otherwise, is the next frontier in neurocognitive research (Malaia, 2017).

The quantitative difference between action and the signal produced with communicative intent (i.e., language, either spoken or signed) is that the linguistic signal contains more information (defined in terms of entropy) across multiple timescales (Malaia, Borneman, & Wilbur, 2016; Singh & Theunissen, 2003). The most recent contribution to the existing evidence on this distinction comes from entropy measures in the visual domain (Borneman, Malaia, & Wilbur, 2018; Bosworth, Bartlett, & Dobkins, 2006; Malaia et al., 2016; Malaia, Borneman, & Wilbur, 2017). A quantitative comparison between the action and sign language signal can be formulated in terms of the amount of motion information in the visual signal, as measured, for example, by optical flow. The optical flow metric describes the change in pixel positions between the two video frames. This method allows to track the motion of all components of the video recording, resulting in an equivalent of a spectrogram for the visual signal (Figure 2a,b). The motion-spectrogram comparison of speech and sign clearly demonstrates that the visual signal in sign language is more variable across time, as well as across frequency bands (i.e., scales). The combination of entropy differences across multiple scales can be mathematically captured by using fractal dimension—an index of signal complexity in the frequency domain that estimates how the variability of the signal changes with the scale at which it is measured. Fractal complexity is a composite measure of entropy across multiple scales. For sign languages, it is higher than for human everyday motion (Figure 2c).

Considering the communicative pressures that shape languages, these results are not surprising. ASL, as any other language, is highly efficient in transferring information in time; there is less to comprehend in action than in language, despite the fact that they are in the same modality. ASL is known to operate on multiple timescales using hands, face, head, and posture as linguistic articulators across different temporal scales (Malaia et al., 2017). For example, the eyebrow furrow scopes



(a) Binned optical flow diagram (visual spectrogram of motion) in a 45-s segment of American Sign Language narrative.
(b) Binned optical flow diagram (visual spectrogram of motion) in a 45-s segment of everyday motion (data from Zacks et al., 2009).
(c) Example normalized spectral density for ASL (red) and everyday motion (blue) show the respective fit according to equation
Signing videos show greater fractal complexit $M(f) = a/f^{\beta}$ across frequencies.

FIGURE 2 ASL and action: comparative variability optical flow spectrograms (a and b) and power law model of the signal across frequencies (c), indicative of information transfer capacity at the observed scale, which is higher for sign language, as compared to action

over the interrogative clause, regardless of how many hand signs are in it; the syntactic structure is put together at a different temporal scale than the lexical word.

So far, this section has focused on global information-theoretic measures of the potential informativity of the (linguistic or nonlinguistic) perceptual signal within and across temporal scales. Other information-theoretic measures make it possible to quantify the local informativity of individual signal units within a particular linguistic level of analysis (phonology, morphology, lexicon, syntax, …). For example, informativity at the morphological, semantic or syntactic levels has been quantified in terms of information-theoretic concepts like *surprisal* and *entropy reduction.* Both metrics can be computed based on corpora[4] and quantify the information load of an incoming stimulus in terms of how strongly it modifies the processor's current model of the world (Frank, 2013; Frank, Otten, Galli, & Vigliocco, 2015; Hale, 2016; Linzen & Jaeger, 2015).

For illustration, let us consider an artificial language. This language has a verb, VERB1, which can only be followed by two different objects, OBJECT1 and OBJECT2 (cf. Figure 3a). The same language has another verb, VERB2, which can appear with nine different objects (OBJECT1 through OBJECT9, cf. Figure 3b). In language use, VERB1 is followed by OBJECT1 in 20% of its occurrences, and by OBJECT2 in the remaining 80% of occurrences. VERB2 is followed by OBJECT1 in 20% of occurrences, but each of the eight remaining options occurs in 10% of cases.

Surprisal (not to be confused with the notion of "surprise"), is essentially a syntagmatic measure: It derives from, and is inversely related to, the forward transitional probability (TP) from one unit (e.g., the verb) to the next (e.g., the object) in the linear sequence of the unfolding linguistic signal. In the artificial language at hand, OBJECT1 has the same surprisal value after VERB1 and VERB2, since the relevant TPs are identical: p(VERB1 + OBJECT1) = 0.20 and p(VERB2 + OBJECT1) = 0.20. On this metric, both instances of OBJECT1 should therefore elicit the same cognitive load. By contrast, OBJECT2 features different surprisal values after VERB1 and VERB2, since TPs are different: p(VERB1 + OBJECT2) = 0.80, whereas p(VERB2 + OBJECT2) = 0.10. This suggests that OBJECT2 should trigger lower cognitive load after VERB1 than VERB2. However, the surprisal metric does not capture one essential fact: After VERB1, OBJECT1 is the paradigmatically dispreferred option, since its only attested competitor, OBJECT2, is four times more frequent. By contrast, after VERB2, OBJECT1 is the paradigmatically preferred option, since it is more likely than any other competitor. Surprisal can thus be said to be agnostic to paradigmatic competition between items that can fill the same slot.

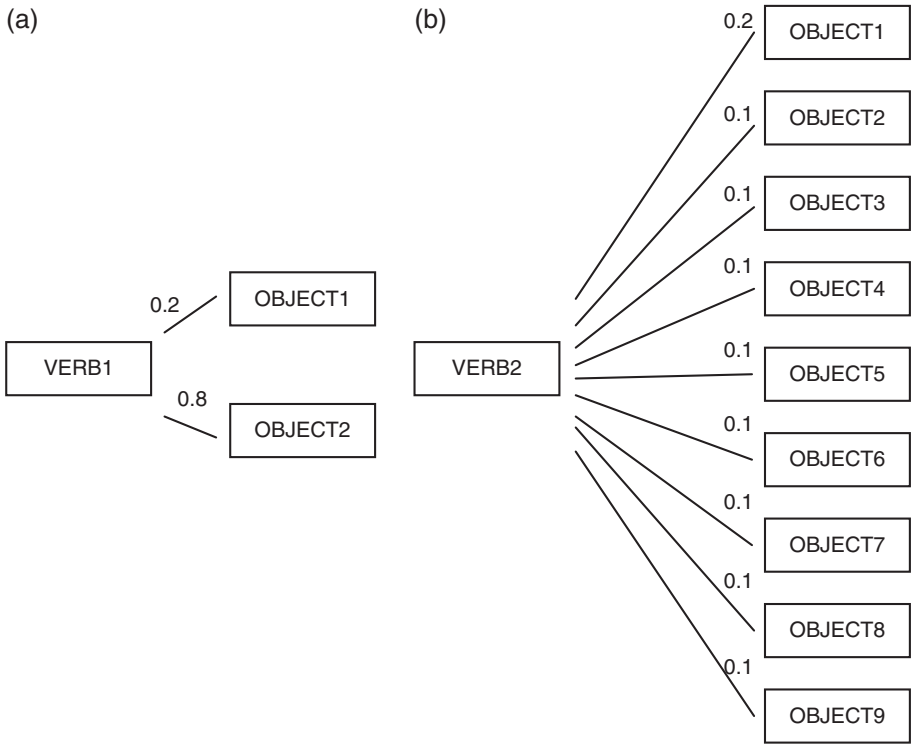


FIGURE 3 Schematic representation illustrating the fact that identical probabilities of syntagmatic co-occurrence (e.g., p(VERB1 + OBJECT1) = 0.20 and p(VERB2 + OBJECT1) = 0.20) need not amount to equal expectedness, since paradigmatic competition between different options (here: objects) will also play a role. Thus, OBJECT1 is the paradigmatically dispreferred competitor after VERB1, while it is the preferred option after VERB2

Entropy reduction is different from surprisal in that it quantifies the extent to which an incoming stimulus reduces prior paradigmatic uncertainty (i.e., entropy) between competing options. In the artificial language under consideration, paradigmatic uncertainty as to the upcoming unit is higher at VERB2 than VERB1, since VERB2 is associated with more potential object continuations and a more balanced probability distribution between these continuations. As a result, any OBJECT after VERB2 will be more uncertainty-reducing (and thus more informative) than any OBJECT after VERB1. However, entropy reduction does not consider the syntagmatic likelihood of individual continuations.

Surprisal and entropy reduction have both been shown to be good predictors of processing cost across different levels of language description, from morphology via the levels of lexical combinatorics up to syntax (Blumenthal-Dramé, 2016b; Blumenthal-Dramé et al., 2017; McConnell & Blumenthal-Dramé, 2018).

A separate family of information-theory-based metrics for action and language include those derived from power laws. In mathematical terms, a power law is a relationship between two parameters of the system (e.g., the number of different lexical items in a language, and the relative frequency of their use), such that a relative change in one quantity results in a proportional relative change in the other quantity. The example of the relationship between a lexical item's ranking on the frequency list and its usage frequency was first formulated as Zipf's law, but the application of the analysis to corpora diachronically and cross-linguistically made it clear that the specific parameters of power-law-based models (such as the fractal complexity parameter, β) reflect not just the properties of the linguistic system at the lexical level, but can be used to track relative changes in syntax from synthetic (high use of morphological modification on lexical items) to analytic (reliance on function words), or to compare the amount of information from various sources in visual communication (Malaia et al., 2016, 2017).

5 | MULTISCALE PREDICTIVE PROCESSING ACROSS DOMAINS

Across all scales of the hierarchy, the comprehension of action and language sequences draws on brain circuits, cognitive representations and controller–predictor modules that strongly overlap with those used for production (Chater et al., 2016; Friston, Mattout, & Kilner, 2011; Pickering & Garrod, 2013; de Wit & Buxbaum, 2017; Wolpert, Doya, & Kawato, 2003). The MSIT framework therefore encompasses representations underlying both comprehension and production processes. Consistent with this claim, recent language and action processing studies have revealed extensive neural activation overlap between the production and comprehension processes within each domain, respectively (Grafton, 2009; Silbert, Honey, Simony, Poeppel, & Hasson, 2014). Furthermore, comprehension studies have provided evidence in support of a cortical processing

hierarchy that is consistent with the functional scales of the MSIT framework (Grafton, 2009; de Heer, Huth, Griffiths, Gallant, & Theunissen, 2017; Yeshurun, Nguyen, & Hasson, 2017).

One key ingredient to the online comprehension of both types of sequences is predictive processing (Huang & Rao, 2011; Lupyan & Clark, 2015; Pickering & Garrod, 2013). Predictive processing in its technical sense can either be sequential (and thus future-directed, e.g., what are likely upcoming units), or top-down (where higher-level schemas and contextual information modulate the interpretation of incoming sensory input). The function of prediction is to restrict interpretation effort for ambiguous input, to support the understanding of noisy input (e.g., unfamiliar accents), to facilitate the integration of incoming information, and to enhance turn-taking and coordination (Huettig, 2015).

Predictions involved in the understanding of familiar, goal-directions actions have been shown to modulate low-level processing, mainly based on studies demonstrating that greater experience with some motor sequence correlates with enhanced prediction capacity and low-level motor excitability while observing it. Thus, professional basketball players are better at predicting the success of basket shots than sports journalists or coaches. Moreover, they show enhanced motor excitability in the predictive observation of basket shots relative to the observation of soccer kicks (Aglioti, Cesari, Romani, & Urgesi, 2008; Costantini, Ambrosini, Cardellicchio, & Sinigaglia, 2014; Elsner, D'Ausilio, Gredebäck, Falck-Ytter, & Fadiga, 2013). In a similar vein, it is widely acknowledged that predictions play a role at all levels of language understanding (Kuperberg & Jaeger, 2016). At the morphological level, a masked visual priming experiment found that lexical decision times to complex words (e.g., *tearless*, *worthless*) correlate with surprisal for the suffix (*−less*), given the base (*tear*, *worth*) (Blumenthal-Dramé et al., 2017). In this experiment, native speakers of English had to decide, as quickly and accurately as possible, whether a string of letters was a possible word of English, after brief (60 ms) exposure to the first morpheme of the word (e.g., *worth—WORTHLESS*, *ticket—TICKETMENT*). This experiment also demonstrated that more surprising base-suffix combinations (e.g., *soft + ish*) yield stronger BOLD activation than less surprising ones (e.g., *doubt + ful*) in regions attributed to language (Bozic, Marslen-Wilson, Stamatakis, Davis, & Tyler, 2007) or general task performance difficulty (Fedorenko, Duncan, & Kanwisher, 2013) (notably bilateral inferior frontal gyri) and to attention process or response conflict (Aarts, Roelofs, & van Turennout, 2009) (superior frontal gyrus, extending into the bilateral medial frontal and cingulate gyri) (cf. Figure 4).

In a reading self-paced reading study, readers were shown to exploit the verb to predict whether a currently incoming sequence will turn out to be a prepositional object construction (*Emma sent John to the doctor*) or a double object construction (*Emma sent John a book*). Remarkably, syntactic entropy reduction was predictive of reading times for structure-disambiguating words (i.e., *a*), whereas surprisal did not have any predictive power (Blumenthal-Dramé, 2018).

While these results are in contrast with those of a range of neurolinguistic and behavioral studies which identified surprisal as a good predictor of lexical unexpectedness (Frank et al., 2015; Smith & Levy, 2013), the differences in findings suggest that higher-level (structural) predictions might depend on more complex and wider-scope metrics than lower-level (lexical) predictions. Within the MSIT framework, the differences among the results of experiments which entail the use of different, yet interdependent scales of linguistic processing (perceptual-syllabic, morphological, lexico-semantic, syntactic, pragmatic) can be straightforwardly accommodated, and appropriate information theory metrics selected for predicting participant behavior based on the properties of the linguistic system.

Parallel findings from action research also suggest cumulative effects of scale on the processing of visual scenes. A neuroimaging study of the fine- and coarse-grained processing of video clips (Zacks, Braver, et al., 2001) showed that the timeline of activation for neural regions attributed to the processing of event boundaries precedes the actual event boundary. This anticipatory activation was observed as early as 5 s before the event boundary in a passive viewing condition, and 15 s prior to it in an active segmentation task condition; the activation for fine-grained boundaries was weaker as compared to coarse-grained ones. The MSIT framework captures that the processing of fine- and coarse-grained boundaries is contingent on each other in top-down and bottom-up processing. It can also accommodate the fact that the entropy flow from lower levels (fine-grained actions, which are lower on the temporal scale) results in a cumulative effect, such that different levels of probability need to be taken into account in the segmentation of coarse-grained events. The higher processing load resulting from multilevel predictive processing at the longer temporal scales for coarse-grained events are then associated with increased metabolic load on the neural regions attributed to predictive processing, as observed in fMRI data.

5.1 | Cues for predictive processing across scales

Action observers and language predictors exploit cues across all scales of the hierarchy (Figure 1) to make predictions. Thus, action observers rely on subtle kinematic information corresponding to the bottom layer of the hierarchy to infer proximal and distal action goals (e.g., the preshaping of the hand can indicate the size of a to-be-grasped target object) (Donnarumma, Dindo, & Pezzulo, 2017). Likewise, language comprehenders draw on fine-grained information, such as anticipatory coarticulation (i.e., the articulatory effect of upcoming sounds on a given sound), but also on morphological and prosodic cues, to

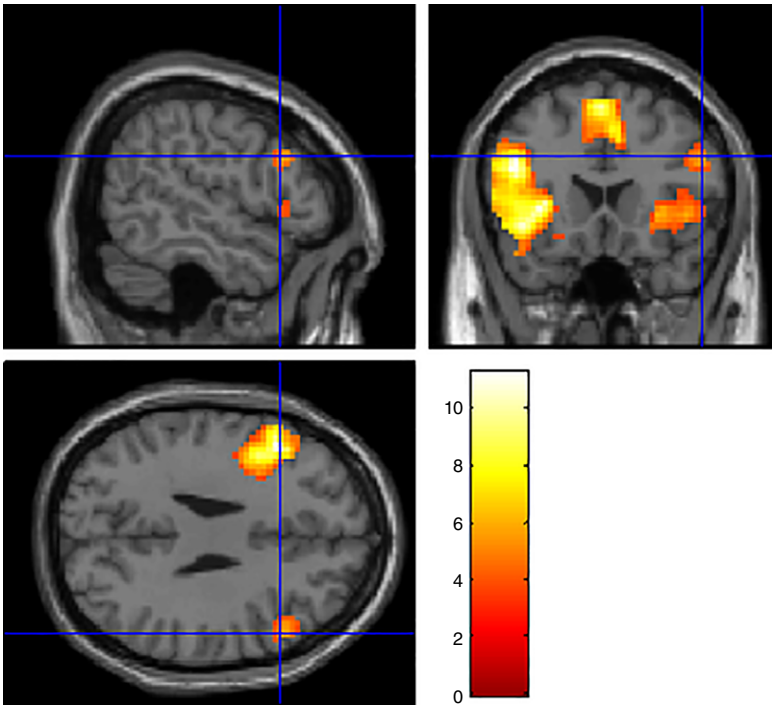


FIGURE 4 fMRI scans showing frontal regions where more surprising base-suffix combinations (e.g., *tear* + *less*) elicit stronger activation than less surprising ones (e.g., *worth* + *less*) in masked visual priming (e.g., tear—TEARLESS vs. worth—WORTHLESS). The positive parametric correlation between BOLD signal and surprisal is projected on sections of the canonical montreal neurological institute (MNI) single-subject template, rendered at a peak-level threshold of $p < 0.001$ (uncorr.). Color bars indicate the range of the relevant voxel-level *t*-values. The crosshairs locate the origin of the voxel of greatest *t*-statistic in a cluster of the right inferior frontal gyrus (*pars opercularis*)

predict upcoming low- and high-level information in spoken and signed languages alike (Henry, Jackson, & Dimidio, 2017; McDonald et al., 2016; McMurray & Jongman, 2016).

At the same time, action observers and language comprehenders also draw on high-level contextual knowledge, which top-down modulates the interpretation of incoming signals at low levels. For example, contextual cues to others' intentions have been shown to modulate corticospinal excitability (which reflects low-level processing at the sensorimotor level) in observers of action sequences, with congruent contexts facilitating and incongruent contexts inhibiting excitability (Amoruso, Finisguerra, & Urgesi, 2016). In language, high-level knowledge such as sentential context can also modulate lower levels, for example, the identification of phonetically ambiguous lexical items. Thus, after a noun-biasing sentence onset such as *Valerie hated the…*, comprehenders will interpret a word whose first phoneme is ambiguous between b and p as a noun (e.g., *pie*) rather than a verb (e.g., *buy*) (Fox & Blumstein, 2016). In a similar vein, knowledge about the talker can affect the perception of speech sounds (McMurray & Jongman, 2016).

A further cue to prediction is statistical knowledge extracted from prior experience. Humans are known to be very proficient at extracting and storing statistical regularities from their motor and linguistic environment, and at exploiting mental schemas representing these regularities to make predictions as to upcoming material (Blumenthal-Dramé, 2016b). Thus, a study on predictive gaze fixations showed that after passive observation, toddlers as young as 8–11 months were able to predict upcoming actions based on statistical relationships extracted from the input, at least when the observed action sequences led up to salient sensory changes (Monroy et al., 2017). Likewise, toddlers of the same age are readily able to extract statistical regularities from spoken syllable sequences after only 2 min of passive exposure (Aslin, 2017).

The human sensitivity to statistical relationships in the input also shows up in experiments where humans are exposed to sequences that violate statistical expectations. Such sequences have been shown to elicit different kinds of surprise responses in the brain. In both action and language understanding, the semantic or conceptual unexpectedness of an incoming stimulus has been shown to correlate with the N400 ERP response in electroencephalography (EEG) (Kaduk et al., 2016). This has been demonstrated, for example, for unexpected conclusions of familiar action sequences (e.g., bringing a pretzel to one's ear vs. to one's mouth) and for semantically incongruous final words of sentences [hearing the word *captains* rather than *dollars* at the offset of the sentence *It was a pleasant surprise to find that the car repair bill was only seventeen…*] (Michel, Kaduk, Ní Choisdealbha, & Reid, 2017; Van Petten & Luka, 2012). Interestingly, 9 months old infants' N400 to unexpected action sequences correlates with their language comprehension abilities at 9 months and their language production abilities at 18 months, suggesting that language and action comprehension are tightly coupled.

In both types of sequences, expectation violations have been claimed to drive efficient online adaptation and learning. In the hierarchical coding framework, each violation of top-down predictions by incoming sensory input results in an error signal which is passed on from lower to higher hierarchy levels and leads to a small update of prior expectations to reduce prediction errors on future input (Christiansen & Chater, 2016).

For example, observers' default expectations about other agents' intentions can be altered through repeated exposure to movies where goals are achieved using biomechanically suboptimal (and thus unexpected) action kinematics. Overt action predictions and mean corticospinal excitability show that observers update their expectations to match them with the input bias. In particular, biased observers show reduced corticospinal activity (and thus reduced motor resonance) compared to observers who have been exposed to a larger number of expectation-congruent movies (Jacquet et al., 2016). Likewise, in language processing, repeated exposure to a priori unexpected syntactic structures triggers expectation adaptation, such that structures that are dispreferred at the beginning of a reading experiment can come to be preferred by its end and vice versa (Fine, Jaeger, Farmer, & Qian, 2013).

Finally, online understanding in both domains has been claimed to follow a "chunk-and-pass procedure" (Christiansen & Chater, 2016). More specifically, this means that as the incoming stream is rapidly segmented into chunks, these chunks are immediately recoded into increasingly abstract units covering successively larger temporal windows, thereby recreating the originally intended hierarchical structure of the language producer/action executor. Thus, in speech comprehension, sequences of phonemes will be grouped and shifted to the syllable level, sequences of syllables will be grouped and shifted to the morpheme level, sequences of morphemes will be grouped and shifted to the lexical level, etc., across all levels of linguistic representation. The same process is supposed to shift motor action information from kinematic up to the intention level (Donnarumma, Maisto, & Pezzulo, 2016; Grafton, 2009). The chunk-and-pass procedure is thought to reduce interference effects between items within low hierarchical levels, and to relieve short-term memory via data compression into less taxing formats. The immediate passing on to higher hierarchical levels can also feed into predictive coding.

This section has reviewed evidence showing that the online comprehension of action and language involves prediction (with predictions relying on analogous cues), and the interaction between different hierarchical levels. In both domains, top-down knowledge guides and contributes to decoding lower-level information. Moreover, expectation violations elicit similar brain responses and contribute to online adaptation and learning in a similar fashion.

6 | CONCLUSION

The MSIT framework provides a continuum within which the low- and high-level processing of the signal (visual or auditory) can be interpreted across different domains, from event observation to language comprehension. The key tenets of the framework are:

1. Both action observation and language perception rely on hierarchical multiscale processing for sense-making.
2. Event observation and language comprehension recruit similar processing algorithms and allow for competence transfer across domains.
3. Potential information contained in the signal can be quantified both locally (i.e., How potentially informative are individual units of the incoming sensory stream?) and globally (i.e., How much information can potentially be extracted from the whole processing stream?) within and across hierarchical scales. Information content of the signal determines engagement of neural processing algorithms in communication across modalities.

The framework gives rise to hypotheses across multiple domains of language study. The predictive power of the MSIT framework can be tested across multiple domains of inquiry, including neurolinguistics, historical linguistics, corpus analysis, language acquisition research, and speech and hearing sciences. Here are a few examples of potential applications of the framework:

- Use of quantitative modeling of information transfer to predict the timelines of language acquisition and learning. The asynchronous acquisition of different analytical levels of language can be modeled in the frequency domain with regard to the complexity of transferred information (Williams et al., 2015).
- Achieving a more fine-grained and systematic understanding of how informativity (as quantified in terms of information-theoretic metrics) at different scales interacts. For example, is there a principled trade-off between scales such that (spoken or signed) languages exhibiting relatively low entropy at low hierarchical scales (e.g., morphology) exhibit higher entropy at some higher scale (e.g., syntax) and vice versa? Does the relative informativity of different scales vary as a function of language change (Bentz, Kiela, Hill, & Buttery, 2014; Chand, Kapper, Mondal, Sur, & Parshad, 2017)?

- Determining which of the abovementioned information transfer metrics is best-suited to predict behavior and online processing of language, and whether the predictivity of metrics depends on the scale and language under consideration.
- Exploring the extent to which information extraction from signal streams in different modalities critically relies on the same brain circuits. This could, for example, be achieved via transcranial magnetic stimulation (TMS) studies exploring whether disrupting information extraction at a given hierarchical scale has analogous processing effects across different modalities (e.g., the processing of motor event schemas vs. linguistic schemas).
- Investigating the relationship between the cognitive capabilities put to use in information processing (e.g., segmentation and hierarchical structuring) and other cognitive processes. Such investigations can take various forms. For example, event segmentation appears to interfere with attention regulation: people are less accurate at detecting probes in a video of ongoing activity at event boundaries than at nonboundaries (Huff, Papenmeier, & Zacks, 2012), suggesting that event boundaries modulate attention even when irrelevant to the task. Likewise, multiple components of executive processing, including attentional regulation as well as the maintenance and manipulation of event representations, have been identified as important to the ability to apply event schemas (Malaia et al., 2009; Malaia & Newman, 2015a, 2015b; Mc Elree, 2006; Nee & Jonides, 2013). It is possible that executive skills predict both event memory and sensitivity to hierarchical and segmental event structures; this suggestion is readily transferable to language as well (Aslin, 2017; Fox & Blumstein, 2016; Monroy et al., 2017). In MSIT framework terms, the flexibility of attentional deployment to information-rich portions of the activity stream appears to undergird effective event identification and the hierarchical organization of event-based models in memory. The interaction of executive processes with signal entropy, as well as action recognition and language networks, requires further investigation using quantitative approaches, such as those laid out within the MSIT framework.

MSIT is a unifying framework for formulating interdisciplinary research that aims at an integrated understanding of human information processing across analytical levels and modalities. It generalizes across research findings from action understanding, sign language processing, and spoken language comprehension work, and encourages the development of quantitative, predictive models of information transfer in a unified domain of perception, action, and cognition.

ACKNOWLEDGMENTS

Authors are grateful to the Freiburg Institute for Advanced Studies for supporting this interdisciplinary work. Both authors contributed equally to this study. Preparation of this paper was funded by Grant #1734938 from the U.S. National Science Foundation and European Union Marie S. Curie FRIAS COFUND Fellowship Programme (FCFP) award to E.M.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

ENDNOTES

[1]As is typical of language systems, the rule does not spread to all verbs in the language—a phenomenon comparable to the production of regular and irregular verbs in English.

[2]*The witness protected by the agent was in danger* (animate, atelic) versus *The mansion seized by…* (inanimate, telic) versus *The witness seized by…* (animate, telic) versus *The mansion protected by…* (inanimate, atelic).

[3]We do not suggest that such inferences could be made based solely on overlapping findings as to regions of neural activity; this would constitute reverse inference. However, the experimental designs of the original studies isolated the same cognitive process in the domains of language and action, which improves confidence in the validity of the inference by increasing the prior probability of the cognitive process in question taking place (Poldrack, 2006).

[4]Large electronic database consisting of text samples (written, or transcribed signed or spoken).

RELATED WIREs ARTICLES

People watching: visual, motor, and social processes in the perception of human movement
The perception of emotion in body expressions
Sound symbolism: the role of word sound in meaning
Statistical methods in language processing

REFERENCES

Aarts, E., Roelofs, A., & van Turennout, M. (2009). Attentional control of task and response in lateral and medial frontal cortex: Brain activity and reaction time distributions. *Neuropsychologia*, *47*(10), 2089–2099. https://doi.org/10.1016/j.neuropsychologia.2009.03.019

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*(9), 1109–1116. https://doi.org/10.1038/nn.2182

Amoruso, L., Finisguerra, A., & Urgesi, C. (2016). Tracking the time course of top-down contextual effects on motor responses during action comprehension. *Journal of Neuroscience*, *36*(46), 11590–11600. https://doi.org/10.1523/JNEUROSCI.4340-15.2016

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in Cognitive Science*, *6*(3), 359–370. https://doi.org/10.1111/tops.12096

Arnold, A., Wing, A. M., & Rotshtein, P. (2017). Building a Lego wall: Sequential action selection. *Journal of Experimental Psychology*, *43*(5), 847–852. https://doi.org/10.1037/xhp0000382

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *WIREs Cognitive Science*, *8*(1–2), e1373. https://doi.org/10.1002/wcs.1373

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychonomic Bulletin & Review*, *23*, 1122–1142. https://doi.org/10.3758/s13423-016-1028-3

Bentz, C., Kiela, D., Hill, F., & Buttery, P. (2014). Zipf's law and the grammar of languages: A quantitative study of old and modern English parallel texts. *Corpus Linguistics and Linguistic Theory*, *10*(2), 175–211.

Blumenthal-Dramé, A. (2012). *Entrenchment in usage-based theories: What corpus data do and do not reveal about the mind*. Berlin, Germany: de Gruyter Mouton.

Blumenthal-Dramé, A. (2016a). Entrenchment from a psycholinguistic and neurolinguistic perspective. In H. J. Schmid (Ed.), *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Boston, Berlin: De Gruyter. https://doi.org/10.1515/9783110341423-007

Blumenthal-Dramé, A. (2016b). What corpus-based cognitive linguistics can and cannot expect from neurolinguistics. *Cognitive Linguistics*, *27*(4), 493–505. https://doi.org/10.1515/cog-2016-0062

Blumenthal-Dramé, A. (2018). The online processing of the English dative alternation: Verb entropy effects. Manuscript submitted for publication.

Blumenthal-Dramé, A., Glauche, V., Bormann, T., Weiller, C., Musso, M., & Kortmann, B. (2017). Frequency and chunking in derived words: A parametric fMRI study. *Journal of Cognitive Neuroscience*, *29*(7), 1162–1177. https://doi.org/10.1162/jocn_a_01120

Boeckx, C. A., & Fujita, K. (2014). Syntax, action, comparative cognitive science, and Darwinian thinking. *Frontiers in Psychology*, *5*, 627. https://doi.org/10.3389/fpsyg.2014.00627

Borneman, J. D., Malaia, E., & Wilbur, R. B. (2018). Motion characterization using optical flow and fractal complexity. *Journal of Electronic Imaging*, *27*(05), 1. https://doi.org/10.1117/1.JEI.27.5.051229

Bosworth, R. G., Bartlett, M. S., & Dobkins, K. R. (2006). Image statistics of American sign language: Comparison with faces and natural scenes. *Journal of the Optical Society of America A*, *23*(9), 2085–2096.

Bozic, M., Marslen-Wilson, W. D., Stamatakis, E. A., Davis, M. H., & Tyler, L. K. (2007). Differentiating morphology, form, and meaning: Neural correlates of morphological complexity. *Journal of Cognitive Neuroscience*, *19*(9), 1464–1475.

Braun, D. A., Mehring, C., & Wolpert, D. M. (2010). Structure learning in action. *Behavioural Brain Research*, *206*(2), 157–165. https://doi.org/10.1016/j.bbr.2009.08.031

Bybee, J. (2003). *Phonology and language use*. Cambridge, England: Cambridge University Press.

Chand, V., Kapper, D., Mondal, S., Sur, S., & Parshad, R. D. (2017). Indian English evolution and focusing visible through power laws. *Language*, *2*(4), 26. https://doi.org/10.3390/languages2040026

Chater, N., McCauley, S. M., & Christiansen, M. H. (2016). Language as skill: Intertwining comprehension and production. *Journal of Memory and Language*, *89*, 244–254. https://doi.org/10.1016/j.jml.2015.11.004

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62. https://doi.org/10.1017/S0140525X1500031X

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204. https://doi.org/10.1017/S0140525X12000477

Clerget, E., Andres, M., & Olivier, E. (2013). Deficit in complex sequence processing after a virtual lesion of left BA45. *PLoS One*, *8*(6), e63722. https://doi.org/10.1371/journal.pone.0063722

Clerget, E., Winderickx, A., Fadiga, L., & Olivier, E. (2009). Role of Broca's area in encoding sequential human actions: A virtual lesion study. *Neuroreport*, *20*(16), 1496–1499. https://doi.org/10.1097/WNR.0b013e3283329be8

Costantini, M., Ambrosini, E., Cardellicchio, P., & Sinigaglia, C. (2014). How your hand drives my eyes. *Social Cognitive and Affective Neuroscience*, *9*(5), 705–711. https://doi.org/10.1093/scan/nst037

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The Hierarchical Cortical Organization of Human Speech Processing. *Journal of Neuroscience*, *37*(27), 6539–6557. https://doi.org/10.1523/JNEUROSCI.3267-16.2017

de Wit, M. M., & Buxbaum, L. J. (2017). Critical motor involvement in prediction of human and non-biological motion trajectories. *Journal of the International Neuropsychological Society*, *23*(2), 171–184. https://doi.org/10.1017/S1355617716001144

Diedrichsen, J., & Kornysheva, K. (2015). Motor skill learning between selection and execution. *Trends in Cognitive Sciences*, *19*(4), 227–233. https://doi.org/10.1016/j.tics.2015.02.003

Donnarumma, F., Dindo, H., & Pezzulo, G. (2017). Sensorimotor coarticulation in the execution and recognition of intentional actions. *Frontiers in Psychology*, *8*, 237. https://doi.org/10.3389/fpsyg.2017.00237

Donnarumma, F., Maisto, D., & Pezzulo, G. (2016). Problem solving as probabilistic inference with subgoaling: Explaining human successes and pitfalls in the tower of Hanoi. *PLoS Computational Biology*, *12*(4), e1004864. https://doi.org/10.1371/journal.pcbi.1004864

Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Computational Biology*, *5*(3), e1000302. https://doi.org/10.1371/journal.pcbi.1000302

Elsner, C., D'Ausilio, A., Gredebäck, G., Falck-Ytter, T., & Fadiga, L. (2013). The motor cortex is causally related to predictive eye movements during action observation. *Neuropsychologia*, *51*(3), 488–492. https://doi.org/10.1016/j.neuropsychologia.2012.12.007

Everaert, M. B. H., Huybregts, M. A. C., Chomsky, N., Berwick, R. C., & Bolhuis, J. J. (2015). Structures, not strings: Linguistics as part of the cognitive sciences. *Trends in Cognitive Sciences*, *19*(12), 729–743. https://doi.org/10.1016/j.tics.2015.09.008

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences*, *1169*(1), 448–458. https://doi.org/10.1111/j.1749-6632.2009.04582.x

Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A. C., Pozzo, T., … Fadiga, L. (2009). Encoding of human action in Broca's area. *Brain*, *132*(7), 1980–1988. https://doi.org/10.1093/brain/awp118

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *110*, 16616–16621. https://doi.org/10.1073/pnas.1315235110

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS One*, *8*(10), e77661.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology*, *61*(6), 825–850. https://doi.org/10.1080/17470210701623605

Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, *1316*(1), 87–104. https://doi.org/10.1111/nyas.12406

Fox, N. P., & Blumstein, S. E. (2016). Top-down effects of syntactic sentential context on phonetic processing. *Journal of Experimental Psychology*, *42*(5), 730–741. https://doi.org/10.1037/a0039965

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*, *5*(3), 475–494. https://doi.org/10.1111/tops.12025

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, *104*(1–2), 137–160. https://doi.org/10.1007/s00422-011-0424-z

Fujimori, A., & Di Sciullo, A. M. (2012). The association of sound with meaning: The case of telicity. In A. M. Di Sciullo (Ed.), *Towards a biolinguistic understanding of grammar: Essays on interfaces* (pp. 141–166). Amsterdam, Netherlands: John Benjamins Publishing.

Garrod, S., Gambi, C., & Pickering, M. J. (2014). Prediction at all levels: Forward model predictions can enhance comprehension. *Language, Cognition and Neuroscience*, *29*(1), 46–48. https://doi.org/10.1080/01690965.2013.852229

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

Goldberg, A. E. (2016). Partial productivity of linguistic constructions: Dynamic categorization and statistical preemption. *Language and Cognition*, *8*(03), 369–390. https://doi.org/10.1017/langcog.2016.17

Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, *23*(4), 959–978. https://doi.org/10.3758/s13423-015-0860-1

Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Annals of the New York Academy of Sciences*, *1156*(1), 97–117. https://doi.org/10.1111/j.1749-6632.2009.04425.x

Grafton, S. T., & Hamilton, A. F. d. C. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*, *26*(4), 590–616. https://doi.org/10.1016/j.humov.2007.05.009

Hale, J. (2016). Information-theoretical complexity metrics. *Language and Linguistics Compass*, *10*(9), 397–412. https://doi.org/10.1111/lnc3.12196

Henry, N., Jackson, C. N., & Dimidio, J. (2017). The role of prosody and explicit instruction in processing instruction. *The Modern Language Journal*, *101*(2), 294–314. https://doi.org/10.1111/modl.12397

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *WIREs Cognitive Science*, *2*(5), 580–593. https://doi.org/10.1002/wcs.142

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

Huff, M., Papenmeier, F., & Zacks, J. M. (2012). Visual target detection is impaired at event boundaries. *Visual Cognition*, *20*(7), 848–864.

Jacquet, P. O., Roy, A. C., Chambon, V., Borghi, A. M., Salemme, R., Farnè, A., & Reilly, K. T. (2016). Changing ideas about others' intentions: Updating prior expectations tunes activity in the human motor system. *Scientific Reports*, *6*, 26995. https://doi.org/10.1038/srep26995

Jeon, H.-A. (2014). Hierarchical processing in the prefrontal cortex in a variety of cognitive domains. *Frontiers in Systems Neuroscience*, *8*, 223. https://doi.org/10.3389/fnsys.2014.00223

Kaduk, K., Bakker, M., Juvrud, J., Gredebäck, G., Westermann, G., Lunn, J., & Reid, V. M. (2016). Semantic processing of actions at 9 months is linked to language proficiency at 9 and 18 months. *Journal of Experimental Child Psychology*, *151*(Suppl. C), 96–108. https://doi.org/10.1016/j.jecp.2016.02.003

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, *15*(8), 352–357. https://doi.org/10.1016/j.tics.2011.06.005

Klima, E. S., Tzeng, O. J., Fok, Y. Y. A., Bellugi, U., Corina, D., & Bettger, J. G. (1999). From sign to script: Effects of linguistic experience on perceptual categorization. *Journal of Chinese Linguistics Monograph Series*, 96–129.

Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron*, *50*(6), 963–974. https://doi.org/10.1016/j.neuron.2006.05.017

Krebs, J., Malaia, E., Wilbur, R. B., & Roehm, D. (2018). Subject preference emerges as cross-modal strategy for linguistic processing. *Brain Research*, *1691*, 105–117. https://doi.org/10.1016/j.brainres.2018.03.029

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79.

Leech, R., & Sharp, D. J. (2014). The role of the posterior cingulate cortex in cognition and disease. *Brain*, *137*(1), 12–32. https://doi.org/10.1093/brain/awt162

Linzen, T., & Jaeger, T. F. (2015). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*, *40*, 1382–1411. https://doi.org/10.1111/cogs.12274

Llanos, F., Alexander, J. M., Stilp, C. E., & Kluender, K. R. (2017). Power spectral entropy as an information-theoretic correlate of manner of articulation in American English. *Journal of the Acoustical Society of America*, *141*(2), EL127–EL133. https://doi.org/10.1121/1.4976109

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language–perception–cognition interface. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226. https://doi.org/10.3389/fpsyg.2013.00226

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2001). Remembering familiar people: The posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience*, *104*(3), 667–676.

Malaia, E. (2004). Event structure and telicity in Russian: An event based analysis for the telicity puzzle in Slavic languages. In *Ohio State University Working Papers in Slavic Studies* (Vol. 4, pp. 87–98). Columbus, OH.

Malaia, E. (2017). Current and future methodologies for quantitative analysis of information transfer in sign language and gesture data. *Behavioral and Brain Sciences*, *40*, e63.

Malaia, E., Borneman, J., & Wilbur, R. B. (2008). Analysis of ASL motion capture data towards identification of verb type. In *Proceedings of the 2008 conference on semantics in text processing* (pp. 155–164). Association for Computational Linguistics.

Malaia, E., Borneman, J. D., & Wilbur, R. B. (2016). Assessment of information content in visual signal: Analysis of optical flow fractal complexity. *Visual Cognition*, *24*(3), 246–251.

Malaia, E., Borneman, J. D., & Wilbur, R. B. (2017). Information transfer capacity of articulators in American sign language. *Language and Speech*, *61*(1), 97–112. https://doi.org/10.1177/0023830917708461

Malaia, E., & Newman, S. (2015a). Neural bases of syntax-semantics interface processing. *Cognitive Neurodynamics*, *9*(3), 317–329. https://doi.org/10.1007/s11571-015-9328-2

Malaia, E., & Newman, S. (2015b). Neural bases of event knowledge and syntax integration in comprehension of complex sentences. *Neurocase*, *21*(6), 753–766. https://doi.org/10.1080/13554794.2014.989859

Malaia, E., Ranaweera, R., Wilbur, R. B., & Talavage, T. M. (2012). Event segmentation in a visual language: Neural bases of processing American sign language predicates. *NeuroImage*, *59*(4), 4094–4101. https://doi.org/10.1016/j.neuroimage.2011.10.034

Malaia, E., Talavage, T. M., & Wilbur, R. B. (2014). Functional connectivity in task-negative network of the deaf: Effects of sign language experience. *PeerJ*, *2*, e446. https://doi.org/10.7717/peerj.446

Malaia, E., & Wilbur, R. B. (2012a). Kinematic signatures of telic and atelic events in ASL predicates. *Language and Speech*, *55*(Pt. 3), 407–421. https://doi.org/10.1177/0023830911422201

Malaia, E., & Wilbur, R. B. (2012b). Telicity expression in the visual modality. In V. Demonte & L. McNally (Eds.), *Telicity, change, and state: A cross-categorial view of event structure* (pp. 122–136). Oxford, England: Oxford University Press.

Malaia, E., & Wilbur, R. B. (2018). Visual and linguistic components of short-term memory: Generalized neural model (GNM) for spoken and sign languages. *Cortex*. https://doi.org/10.1016/j.cortex.2018.05.020

Malaia, E., Wilbur, R. B., & Di Sciullo, A. M. (2012). What sign languages show. In A. M. Di Sciullo (Ed.), *Towards a biolinguistic understanding of grammar: Essays on interfaces* (pp. 265–275). Amsterdam, Netherlands: John Benjamins Publishing.

Malaia, E., Wilbur, R. B., & Milkovic, M. (2013). Kinematic parameters of signed verbs. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1677–1688. https://doi.org/10.1044/1092-4388(2013/12-0257)

Malaia, E., Wilbur, R. B., & Weber-Fox, C. (2009). ERP evidence for telicity effects on syntactic processing in garden-path sentences. *Brain and Language*, *108*(3), 145–158. https://doi.org/10.1016/j.bandl.2008.09.003

Malaia, E., Wilbur, R. B., & Weber-Fox, C. (2012). Effects of verbal event structure on online thematic role assignment. *Journal of Psycholinguistic Research*, *41*(5), 323–345. https://doi.org/10.1007/s10936-011-9195-x

Malaia, E., Wilbur, R. B., & Weber-Fox, C. (2013). Event end-point primes the undergoer argument: Neurobiological bases of event structure processing. In B. Arsenijević, B. Gehrke & R. Marín (Eds.), *Studies in the composition and decomposition of event predicates* (pp. 231–248). Dordrecht, Netherlands: Springer.

Martins, M. D., Martins, I. P., & Fitch, W. T. (2016). A novel approach to investigate recursion and iteration in visual hierarchical processing. *Behavior Research Methods*, *48*(4), 1421–1442. https://doi.org/10.3758/s13428-015-0657-1

Mc Elree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 46, pp. 155–201). San Diego: Academic Press.

McConnell, K., & Blumenthal-Dramé, A. (2018). The online processing of collocations: Effects of task and corpus-derived association scores.

McDonald, J., Wolfe, R., Wilbur, R. B., Moncrief, R., Malaia, E., Fujimoto, S., … Stec, J. (2016). A new tool to facilitate prosodic analysis of motion capture data and a data-driven technique for the improvement of avatar motion. In *Proceedings of Language Resources and Evaluation Conference* (*LREC*) (pp. 153–159).

McMurray, B., & Jongman, A. (2016). What comes after/f/? Prediction in speech derives from data-explanatory processes. *Psychological Science*, *27*(1), 43–52.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Michel, C., Kaduk, K., Ní Choisdealbha, Á., & Reid, V. M. (2017). Event-related potentials discriminate familiar and unusual goal outcomes in 5-month-olds and adults. *Developmental Psychology*, *53*(10), 1833–1843. https://doi.org/10.1037/dev0000376

Milkovic, M., & Malaia, E. (2010). *Event visibility in Croatian sign language: Separating aspect and aktionsart.* Poster session presented at the tenth meeting of Theoretical Issues in Sign Language Research (pp. 165–166).

Monroy, C. D., Gerson, S. A., Domínguez-Martínez, E., Kaduk, K., Hunnius, S., & Reid, V. (2017). Sensitivity to structure in action sequences: An infant event-related potential study. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2017.05.007

Moro, A. (2014a). On the similarity between syntax and actions. *Trends in Cognitive Sciences*, *18*(3), 109–110. https://doi.org/10.1016/j.tics.2013.11.006

Moro, A. (2014b). Response to Pulvermüller: The syntax of actions and other metaphors. *Trends in Cognitive Sciences*, *18*(5), 221. https://doi.org/10.1016/j.tics.2014.01.012

Nee, D. E., & Jonides, J. (2013). Neural evidence for a 3-state model of visual short-term memory. *NeuroImage*, *74*, 1–11.

Nielsen, F. A., Balslev, D., & Hansen, L. K. (2005). Mining the posterior cingulate: Segregation between memory and pain components. *NeuroImage*, *27*(3), 520–532. https://doi.org/10.1016/j.neuroimage.2005.04.034

Nuttall, H. E., Kennedy-Higgins, D., Devlin, J. T., & Adank, P. (2017). The role of hearing ability and speech distortion in the facilitation of articulatory motor cortex. *Neuropsychologia*, *94*, 13–22. https://doi.org/10.1016/j.neuropsychologia.2016.11.016

Pastra, K., & Aloimonos, Y. (2012). The minimalist grammar of action. *Philosophical Transactions: Biological Sciences*, *367*(1585), 103–117.

Philipp, M., Graf, T., Kretzschmar, F., & Primus, B. (2017). Beyond verb meaning: Experimental evidence for incremental processing of semantic roles and event structure. *Frontiers in Psychology*, *8*, 1806. https://doi.org/10.3389/fpsyg.2017.01806

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347. https://doi.org/10.1017/S0140525X12001495

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, *17*(9), 458–470. https://doi.org/10.1016/j.tics.2013.06.004

Pulvermüller, F. (2014). The syntax of action. *Trends in Cognitive Sciences*, *18*(5), 219–220. https://doi.org/10.1016/j.tics.2014.01.001

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360. https://doi.org/10.1038/nrn2811

Ranganath, C., Johnson, M. K., & D'Esposito, M. (2003). Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia*, *41*(3), 378–389.

Romagno, D., Rota, G., Ricciardi, E., & Pietrini, P. (2012). Where the brain appreciates the final state of an event: The neural correlates of telicity. *Brain and Language*, *123*(1), 68–74.

Rudge, P., & Warrington, E. K. (1991). Selective impairment of memory and visual perception in splenial tumours. *Brain: A Journal of Neurology*, *114*(Pt. 1B), 349–360.

Shannon, B. J., & Buckner, R. L. (2004). Functional-anatomic correlates of memory retrieval that suggest nontraditional processing roles for multiple distinct regions within posterior parietal cortex. *Journal of Neuroscience*, *24*(45), 10084–10092. https://doi.org/10.1523/JNEUROSCI.2625-04.2004

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, *111*(43), E4687–E4696. https://doi.org/10.1073/pnas.1323812111

Singh, N. C., & Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *Journal of the Acoustical Society of America*, *114*(6), 3394–3411. https://doi.org/10.1121/1.1624067

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, *164*(Suppl. C, 77–105. https://doi.org/10.1016/j.bandl.2016.10.004

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Son, M.-J., & Cole, P. (2008). An event-based account of-kan constructions in standard Indonesian. *Language*, *84*(1), 120–160.

Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(4), 335–345.

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science*, *18*(5), 449–455.

Stilp, C. E., & Kluender, K. R. (2010). Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility. *Proceedings of the National Academy of Sciences*, *107*(27), 12387–12392. https://doi.org/10.1073/pnas.0913625107

Stilp, C. E., & Kluender, K. R. (2016). Stimulus statistics change sounds from near-indiscriminable to hyperdiscriminable. *PLoS One*, *11*(8), e0161001. https://doi.org/10.1371/journal.pone.0161001

Stilp, C. E., Rogers, T. T., & Kluender, K. R. (2010). Rapid efficient coding of correlated complex acoustic properties. *Proceedings of the National Academy of Sciences*, *107*(50), 21914–21919. https://doi.org/10.1073/pnas.1009020107

Stilp, C. E., Kiefte, M., Alexander, J. M., & Kluender, K. R. (2010). Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences. *Journal of the Acoustical Society of America*, *128*(4), 2112–2126. https://doi.org/10.1121/1.3483719

Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M., & Pfau, R. (2015). Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences*, *112*(19), 5968–5973.

Tettamanti, M., & Moro, A. (2012). Can syntax appear in a mirror (system)? *Cortex*, *48*(7), 923–935. https://doi.org/10.1016/j.cortex.2011.05.020

Tettamanti, M., & Weniger, D. (2006). Broca's area: A supramodal hierarchical processor? *Cortex*, *42*(4), 491–494. https://doi.org/10.1016/S0010-9452(08)70384-8

Thompson, J. J., McColeman, C. M., Stepanova, E. R., & Blair, M. R. (2017). Using video game telemetry data to research motor chunking, action latencies, and complex cognitive-motor skill learning. *Topics in Cognitive Science*, *9*(2), 467–484. https://doi.org/10.1111/tops.12254

Valenstein, E., Bowers, D., Verfaellie, M., Heilman, K. M., Day, A., & Watson, R. T. (1987). Retrosplenial amnesia. *Brain: A Journal of Neurology*, *110*(Pt. 6), 1631–1646.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*(9), 445–453. https://doi.org/10.1016/j.tics.2005.07.001

Wilbur, R. B. (2009). Productive reduplication in a fundamentally monosyllabic language. *Language Sciences*, *31*(2), 325–342. https://doi.org/10.1016/j.langsci.2008.12.017

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language*, *101*(3), 278–289. https://doi.org/10.1016/j.bandl.2007.03.004

Williams, J. R., Lessard, P. R., Desu, S., Clark, E. M., Bagrow, J. P., Danforth, C. M., & Dodds, P. S. (2015). Zipf's law holds for phrases, not words. *Scientific Reports*, *5*, 12209. https://doi.org/10.1038/srep12209

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London B*, *358*(1431), 593–602.

Yeshurun, Y., Nguyen, M., & Hasson, U. (2017). Amplification of local changes along the timescale processing hierarchy. *Proceedings of the National Academy of Sciences*, *114*(35), 9475–9480.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*(6), 979–1008.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., … Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*(6), 651–655.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293.

Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain Research*, *1076*(1), 150–162. https://doi.org/10.1016/j.brainres.2005.12.122

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29–58.

How to cite this article: Blumenthal-Dramé A, Malaia E. Shared neural and cognitive mechanisms in action and language: The multiscale information transfer framework. *WIREs Cogn Sci*. 2018;e1484. https://doi.org/10.1002/wcs.1484