Prosodic Rank Theory: on the formalisation of prosodic events

Dafydd Gibbon

1 Prosody – an elusive feature of speech

1.1 The argument

Prosody is not easy to characterise or delimit in relation to other functional patterns of speech, as definitional conflicts over the decades have shown. In the present contribution,[1] it is argued that a comprehensive, event-based semiotic characterisation of prosody is in principle possible, and a systematic approach to the formal modelling of prosodic events is presented, *Prosodic Rank Theory*, which concentrates mainly on temporal relations in intonation and tone systems. The methodological focus is on the use of formal languages to model temporal sequences within the framework of an event-based theory of prosody as a submodality of the voice-to-ear communicative modality.

A subsidiary argument is a critique of conventional linguistic descriptions, not only in the prosodic domain, which are essentially collections of isolated rules or rule-sets with different application domains, whose mutual dependencies are not explicitly formulated in

1 Some of the material in this paper has figured in and benefited from reviews of previous publications, comments on presentations and many discussions with students and colleagues. Parts have already been discussed with Jerzy Bańczerowski in the light of his work in phonology, axiomatic methods and formal semantics during various linguistic conferences and workshops in Poznań. The invitation to contribute to his *Festschrift* is not only a personal pleasure and honour, but gives me the opportunity to gather together in a compact form some of the main ideas in my previous work, and to put them into a broader context. The paper presupposes a certain familiarity with principles of computational phonology and with the main aspects of prosody. Acknowledgments are due to many friends, colleagues and students, particularly to Julie Berndsen and Petra Wagner, but first and foremost to Wiktor Jassem for countless fruitful discussions.

detail and are therefore not clear. There exist rule organising concepts such as *intrinsic ordering*, *extrinsic ordering*, the *phonological cycle*, the *ordered levels* of lexical phonology, the deep and surface *levels* of Chomskyan grammars, the *ranks* of functional linguistics. But linguistic descriptions are very complex, and thorough checking of all compositional structural dependencies is too time consuming for the working hours of an average linguist. Fortunately, many insights about when a description is sound and complete, and about efficient processing, have been introduced into theoretical and descriptive linguistics by formal and computational linguistics (for an overview, cf. MITKOV 2003). The present study seeks to benefit from these increases in efficiency in checking the validity of grammars.

In particular, it is argued in this context that temporal sequencing properties of prosodic patterns are adequately represented by Type III (regular) grammars and finite state machines (FSMs), the most basic and efficiently computable kinds of formal grammar. The typology of temporal sequencing in prosodic systems varies greatly, with functionalities ranging from the lexical domain (from phonemic through morphemic and morphosyntactic structures, and with forms ranging from tones through pitch accents and pitch realisation of stress accents) to semantic phrasal and pragmatic discoursal intonation domains. The core of a new framework, *Prosodic Rank Theory*, is introduced in order to explain prosodic patterning as the phonetic interpretation of structural categories at different locutionary ranks from phoneme to discourse. *Prosodic Rank Theory* generalises over and replaces older concepts of a more restrictive autonomously defined strictly layered Prosodic Hierarchy.

The main focus in this contribution is on modelling tonal forms with phonological and morphological functions at lexical ranks, though the neighbouring domains of intonation and rhythm are also touched on. A panacea and a complete description are not given, but the *gestalt* of a comprehensive theory is developed.

1.2 The structure

This contribution is not an empirical study in itself, though it relies on empirical results from previous work. If there were a discipline of the philosophy of prosody, it would belong there: the innovative features are in the positioning of prosodic modelling in a more general context. The study follows the following specific line of development. In the second section, prerequisites for a theory of prosody are outlined on the basis of a semiotic approach to defining the events which determine modalities and submodalities in space, time and functionality, and time is picked out as the central characterising category. The third section provides an overview of regular languages as models for rhythmic and tonal temporal patterns, in which regular grammars are interpreted by finite automata as operational models. In the fourth section *Prosodic Rank Theory* is introduced, and in the fifth and concluding section, the results are summarised and a perspective for further development of the approach is outlined.

2 Time, space, events and time: Dual Interpretation Theory

2.1 The semiotic background

The core ontology on which this triadic model of composition and dyadic interpretation is based is expressed as the triple <*category*, *object*, *media*>. The semiotic relation itself is dyadic, expressed traditionally as the Saussurean relation between *signifié* and *significant*. In

this *Dual Interpretation Theory (DIT)* framework, the semiotic relation is represented by a model consisting of a dyad of *interpretation functions* rather than as a pair of *domains* as in traditional approaches:

I_{object}: *category* → *object*

I_{media}: *category* → *media*.

Both the *object semantic interpretation* function and the *media semantic interpretation* function represent denotational 'semantic' models in the the formal sense, and generalise the model-theoretic approach which underlies the logical concept of denotational semantics and the Chomskyan concepts of *semantic interpretation* and *phonetic interpretation*. The idea of 'phonetics as semantics' might sound odd from a conventional linguistic point of view, but formally it is clear, and indeed it is a basic principle of Generative Phonology[2] (not often recognised even by its adherents).

The structure of the core ontology of the present semiotic approach is visualised in simplified form in Figure 1. The key concept underlying the model which Figure 1 illustrates is the *world* in which the *communicator* and the communication *media* are located. The communicator organises the structure – composition and interpretation – of signs, and for each instance of a communicator, there is an instance of the entire model. The communicator is is not a monad, if he is lucky, and his world overlaps with that of other communicators: strongly, if the communicators share, first, the same environment, second, the same language, third the same culture, and fourth, the same experiences, but weakly if any of these four conditions are not met. Clearly these conditions can never be fully met: communicators' environments, lanuages, cultures, experienes overlap, but are never identical: homogeneity assumptions about the world are perhaps heuristic necessities, but they are far from the truth.

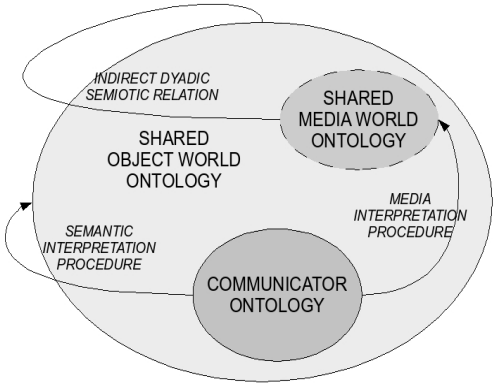

Figure 1: Dyadic semiotic ontology: media semantics as a subset of conventional semantics.

The communicative categories which the communicator generates may be simple or complex, opaque or compositional, and are interpreted by functions mapping them on the one hand into the shared object world, and on the other into the shared media world of a common language, gestural patterning and pictorial conventions. It is this pair of functions which constitutes the semiotic relation; thus the model is triadic, not dyadic, though the functions on their own constitute a dyadic sub-model.

2 "Observe that the interpretive semantic rules must apply in accordance with essentially the same principle as the one stated here for phonological rules" (CHOMSKY & HALLE 1968:20, fn. 7).

The inclusion of the media world as a subset of the shared world explicates the reflexive metalingual[3] functions of language (JAKOBSON 1960). The etalingual functions are both *metalinguistic* ('talking about language') in the conventional sense, but also *metalocutionary* (more specifically: metadeictic), as in prosodic pointing to locutionary constituents (*The CAT, not the HAT*) and in the configurative functions of prosody in speech (GIBBON 1983). The configurative functions include the demarcative functions of boundary tones, and the positioning of lexical, phrasal and focal accents, for which *Prosodic Rank Theory*, an elaboration of *Dual Interpretation Theory*, is introduced.

2.2 Space, time and events

The argumentation of the present study is based initially on the *means* rather than the *functions* of communication.

The means of human communication are conventionally divided into three main modalities:

1. *voice-to-ear* (vocal speech, paralinguistic snorts, grunts, speech surrogates such as whistling, and borderline sounds such as clicks, which figure as phonemes in Khoisan and some Nguni languages, but paralinguistically in other languages, e.g. the 'tut-tut' or 'tst tst' reduplicated click in English, meaning *I disapprove*.),
2. *gesture-to-eye* (signing, lip-reading, conversational limb gestures, posture, facial expression, gaze; operation of sign artefacts),
3. *gesture-to-ear&eye* (clapping, finger snapping, stamping, foot-tapping limb gestures; some vocal gestures such as lip-movements; operation of sign artefacts).

Table 1: Basic classification of communicative gestures in terms of a classification of communication modalities as output-input channels.

Output means	*Input means*	
	ear	*eye*
voice	speech (including both phonemic and paralinguistic vocal clicks), grunts, sniffs, snorts	signing, lip-reading
limbs	clapping, finger snapping, stamping, foot-tapping limb gestures; some vocal gestures such as lip-movements overlap	signing, lip-reading, conversational limb gestures, posture, facial expression, gaze
instrument	tapping, knocking, rattles, whistles; musical instruments; electronic media	writing, inscriptions, pictographic, indexical and symbolic artefacts; visual art; electronic media

The voice-to-ear and gesture-to-ear modalities overlap both in functional and in gestural structure. In contemporary theories both modalities are seen as special cases of an acoustic transduction modality: the speech-producing gestures of the vocal tract, as modelled in Articulatory Phonology and in articulatory phonetics, are a subset of the entire set of gestures. The gesture-to-eye modality overlaps functionally with the gesture-to-ear modality, but

3 The term is often misquoted as 'metalinguistic', but Jakobson uses the term 'metalingual' advisedly, rather than 'metalinguistic', for language functions relating to the functional constitutive factor 'code'.

formally it only overlaps in respect of production but not perception. Consequently, the voice-to-ear and gesture-to-ear modalities are actually special cases of a *gesture-to*-ear auditory transduction modality, the speech-producing gestures of the vocal tract, as modelled in Articulatory Phonology and in articulatory phonetics, being a subset of the entire set of gestures.

Prosody is evidently a feature of the voice-to-ear modality[4], and will be defined here as a distinct *voice-to-ear submodality* of the voice-to-ear modality because of its partial autonomy from the *locutionary submodality* of words, phrases, sentences, texts, dialogue. The gestures in each modality and submodality are *simple events* which originate in specific articulators or *complex events* in which co-articulation occurs, which have both *spatial* and *temporal* distinguishing features. The main focus is on modelling temporal relations between gestures as articulatory events taking place in space and time. A preliminary outline of the event-based ontology underlying the present approach is drawn in GIBBON (2006). Both modalities and submodalities are defined semiotically on the basis of both forms and functions, not in terms of forms, i.e. communicative means, alone. For this purpose P*rosodic Rank Theory* is introduced later in this study.

2.3 Events

The central prosodic ontological categories in the present discussion are *time* and *event*. In anthropology and in linguistic semantics, various models of time have been identified in different cultural systems: linear time, non-linear time, embedded time, cyclical time, calendar time, astronomical time, subjective time, indexical time, differential time, and many others. These concepts of time are all, perhaps surprisingly, not only relevant for object-semantic interpretation, but also for media-semantic interpretation. For example, the calendrical organisation of institutional and mass media communication at discourse level does not necessarily *denote* temporal categories but it is *structured* and *positioned* with reference to temporal categories. Certain utterances both denote the time of day and are located at a time of day (*Good morning!*). Others are located at various times of day but do not denote a time of day (*Hello!*). However, the utterances themselves are necessarily located indexically at particular times. For present purposes in prosody, sequential and parallel events in linear time are the central explicanda, but cyclical time is relevant in the explication of iterative rhythmic and tonal systems.

An approach to 'natural' phonology based on events rather than on abstract atemporal categories such as segments, autosegments, abstract duration or concatenation, is *Event Phonology*, developed by BIRD & KLEIN (1990) as a formalisation of Autosegmental Phonology (cf. GOLDSMITH 1990; LADD 1996) in a declarative logic framework. Event Phonology is based on the Event Logic of Johan van Benthem (VAN BENTHEM 1988), which was developed in the context of logical semantics and the semantic interpretation of sentences. Event Phonology characterises phonology as a domain which can be taken as a formal 'semantic interpretation' of linguistic categories, in the sense of model-theoretic semantics, into the phonetic domain. In the present context, this concept of interpretation is generalised, and extended to the domain of prosody as a whole.

An *event* in the sense of van Benthem and of Bird and Klein is characterised as a pair <*property*, *interval*>, where a property is understood in general terms as a category, specified feature, or attribute-value pair. Events are therefore time functions. The property trajectory

4 Though the term is sometimes used metaphorically in the visual domain, as in sign language 'phonology' and 'prosody', and in descriptions of punctuation and layout as the prosody of writing.

during the interval may be a static or dynamically changing function, and the interval may be expressed as a holistic category or as a pair of points; in the extreme case the interval may be a single point or instant (cf. discussions in CARSON-BERNDSEN 1998 and WAGNER 1998). Events enter into temporal relations with each other: intervals may *precede* each other and *overlap* each other.

In Event Logic and Event Phonology a further relation of *inclusion* is also defined in order to express hierarchies. It is not obvious that inclusion is required as a primitive relation in the present context from a logical point of view. The following condition can be formulated for inclusion of an event, where an event E_2 is overlapped at both ends (where *p*, *o*, and *i* represent the relations of precedence, overlap and inclusion):

IF $p(E_1,E_2)$ & $p(E_2,E_3)$ & $o(E_4,E_1)$ & $o(E_4,E_2)$ & $o(E_4,E_3)$ THEN $i(E_4,E_2)$

It is also not clear whether inclusion is necessary on the grounds that – intuitively – inclusion is required in temporal models for general context-free languages, but not for the more restricted class of regular languages. For cases where the event is initial in the overlapped sequence, it can be stated that there is no event which precedes E_2 and is overlapped by E_4, and in the case of a final event in the overlapped sequence, there is no event which E_2 precedes and is overlapped by E_4. These conditions assume that there are no gaps between events, but does not refer to such gaps.

Prosodic events are specific types of event: a prosodic event is characterised as a pair <*contour*, *interval*>, or more concretely as a pair <*contour*, *length*>, whereby the minimal contour is a property such as a *pitch target*, and the minimal length is a point.

2.4 Time

The *time* category is not homogeneous. In the context of architectures for spoken language processing systems, Gibbon (1992) distinguished between three *Time Types* to which this event concept can be applied. These time types represent different levels of temporal abstraction:

1. *Categorial Time*: contrastive phonological categories such as *duration*, which have temporal interpretations but are actually not specified for temporal properties, and combinatorial operations such as *concatenation* (analogous to the precedence relation) and *unification* of feature bundles (analogous to the overlap relation), which also do not have an explicit termporal interpretation. At this level of abstraction, categories and operations are structural categories, and require the assignment of temporal properties by phonetic interpretation. Whether there are additional morphophonemic constraints on 'abstractness' or phonetic constraints on 'naturalness' of categorial time is not an issue here: structurally and functionally, the Categorial Time Type is still valid.
2. *Relative Time*: temporal relations in phonology and prosody with explicit categories based on temporal organisation such as the syllable and the foot, and temporally relevant interval-defining operations such as autosegmental association, blocking, spreading. At this level of abstraction, temporal relations matter, but it is of no immediate theoretical concern at this level, for instance, whether syllables last 200 ms or 200 years: this is the domain of the Absolute Time Type. The Relative Time Type provides the first stage of media-semantic interpretation for the Categorial Time Type, and is the level at which Event Phonology is located.
3. *Absolute Time*: measurable temporal quantities in phonetics, with numerical interval duration measures. In Absolute Time, the notions of linear and non-linear time require

further specification: temporal compression and expansion, and asymptotic temporal functions require attention. This is the physical and behavioural level at which signal processing operates in phonetics and speech technology. The Absolute Time Type provides a further level of interpretation for the Categorial Time Type via the Relative Time Type. The Absolute Time Type is structured further into temporal subdomains types, corresponding approximately to the three 'prosody' types of TILLMANN (1980). By default, each of these temporal domains relate to a rank in the rank hierarchy of linguistic units in terms of which prosodic functionality is characterised; rank-shifts can take place, depending on pragmatic and phonostylistic factors. The Absolute Time Type subdomains are interpretable functionally and temporally:

phrase:	> 1000 ms	A-prosody (intonation/rhythm group rank)
word:	≈ 700 ms	B-prosody (foot rank)
morpheme:	≈ 250 ms	?-prosody (syllable rank)
segment:	≈ 100 ms	C-prosody (phoneme/allophone rank)

Kornai gave the catchy names of 'Rubber time' and 'Clock time' to the levels of Relative Time and Absolute time.[5] A further temporal concept, 'Cloud Time', may be introduced in order to distinguish between the partitioned clock time of empirical measurements and the analogue time flow of common sense time, to misappropriate Popper's 'clouds and clocks' terminology (POPPER 1965).

Carson-Berndsen's *Time Map Phonology* (CARSON-BERNDSEN 1998) demonstrates for German syllables how the three Time Types are mapped into each other in a composite interpretation of Categorial Time via Relative Time into Absolute Time. Time Map Phonology provides a formal operational model for event and Time Type mapping in the form of a cascade of finite state transducers with feature unification as the pattern matching criterion for transitions. In this approach, at the phonological level an event is a pair of a property and an interval $<P, I>$. At the phonetic level, an event is a quadruple $<P, t_S, t_F, C>$, providing information on event-type (property), start of interval, end of interval and confidence value (CARSON-BERNDSEN & GIBBON 1992). Wagner (WAGNER 1998) takes a related view, but proposes a point-based ontology for both phonology and phonetic levels, and indicates a strategy for defining a syntax-prosody interface with both unification and finite state based models; cf. also (WAGNER 2002) for finite state modelling of stress patterns.

The mereological approach taken in Batóg's Axiomatic Phonology (BATÓG 1967) is the first explicit contribution to the formalisation of temporally interpretable part-whole structures in segmental phonology, and is located at the Relative Time Type level. In conventional phonological terms, 'underlying' structural models of phonology, morphology and syntax (and rank levels modelling larger units, in functional models of linguistics) are concerned with Categorial Time and to some extent, Relative Time. The phonetic interpretation of underlying categories, most explicitly in the case of prosodic phonologies such as Autosegmental-Metrical Phonology, Beats-and-Binding Phonology in the Natural Linguistics framework (DZIUBALSKA-KOŁACZYK 2002), Government Phonology (KAYE 1989), introduces a similar explicit abstraction level of Relative Time to the level which is modelled in Event Phonology. In Articulatory Phonology (BROWMAN & GOLDSTEIN 1992), and in phonetic theories of speech production, transmission and perception, and of course in speech technology applications, both Relative Time Types and Absolute Time Types are required.

5 Personal communication during discussions at the ESSLI Summer School, U Essex, 1998.

2.5 Space

Although the aspect of spatial relations in Relative Space and Absolute Space will not be elaborated in the present study, for a comprehensive theory of prosody it is necessary to proceed beyond Event Phonology and Time Map Phonology, and to take into consideration that events have a *location* as well as a *time*. The gestural events which operate in each modality are distinguished spatially by their location on the body and by their position in relation to some reference point such as a position of rest, or the torso, as well as by their complex indexical spatial relation to interlocutors and to the physical environment. Clearly spatial properties (and other properties of the gesturing organs such as weight, muscular endowment, innervation) also influence the temporal properties of gestures, but will not be formulated in more detail here. It is sufficient for the moment to note that gestures in the voice-to-ear modality are located in the vocal tract, others are located elsewhere on the face and the rest of the body. There are overlaps, such as lip and jaw movement and communicative sounds of snapping, clapping, stamping and tapping which are articulated by hands and legs, which are involved in both modalities. A model of spatiality as well as of temporality is an essential part of prosodic theory.

Spatially arranged clusters of gesture articulators, which will be referred to as *gesture clusters*, are explicated at the Relative Time Type level in the *Feature Geometry* development of Clements' original Autosegmental Phonology framework (CLEMENTS 1981). Perhaps the most popular of these clusters in prosody studies are the gestures of the larynx associated with intonation, tone, pitch accent and the realisation of stress. But the larynx and its parts are also the source of other gestures: glottal stops and fricatives, and vocalisation types such as modal, breathy and creaky voice. Conversely, other gestures are also associated with prosody in different languages and different linguistic frameworks: in the Prosodic Phonologies (such as Firthian Phonology and Autosegmental Phonology), nasal, lateral, labial and other prosodies, in the latter case referred to as 'melodies', are identified.

The concept of event introduced in Event Phonology operates in a purely temporal world. A purely temporal world constitutes a very strong abstraction away from the spatio-temporal world of common-sense characterisation of individuals, as Strawson discusses (STRAWSON 1959), and from the spatially related objects (in this case: articulators) which generate prosodic events. For a full formalisation of the spatio-temporal properties of gestural communication, a leaf can be taken, for example, out of the book of *Situation Semantics* (BARWISE & PERRY 1983), whose ontology includes spatial and temporal locations as well as individuals and situations.

After a brief excursus into the spatial domain which is required for a full theory of prosody, the present argument returns to the formal modelling of events and their temporal relations.

3 Regular models of temporal relations in prosody

3.1 Modelling events and temporal relations

The direction taken in this section is to enquire how simple and complex events may be modelled, selecting utterances as time functions of sequential events for attention, rather than

parallel events and the event overlap relation. Generalising a commonplace claim from logical and generative theories of language, the speech of a given communicator is a potentially infinite set of utterances *U*, implying that either there is a set of infinite event vocabularies *I* from which the constituent events of each utterance $u \in U$ are taken, or that there is no bound on the length of a $u \in U$. It was pointed out in a famous *Linguist List* discussion in 1991[6] that assumptions of this kind can be called into question. The actual speech of any given communicator is evidently a finite set because of the temporal constraints on mortality, but it may still be asked whether speech as a set of utterances in principle a finite set, or an unbounded family of finite sets, or a set containing utterances of infinite length, that is, taking infinite time or (for written inscriptions, space)? The latter assumption may seem bizarre, but the traditional claim in linguistics that there is no bound on the maximal length of sentences has the obvious corollary that sentences may have infinite length. To exclude this, an additional assumption of finiteness is necessary, namely that the length of inscriptions (i.e. their spatial extent) and utterances (i.e. their temporal extent) is bounded. This is evidently a nice theoretical point, but not an empirical question.

Given a modelling convention with infinite sets of finitely bounded utterances, the simplest formal model for sequential temporal relations is the finite state automaton, with simple events modelled by the transitions between states, and complex events modelled by paths through the automaton, and repeated sequences modelled by iterative loops in the automaton and linear recursions in the corresponding formal grammar. An operational model for the temporal relation of overlap is provided by multi-tape finite state transducers (KAPLAN & KAY 1994). Equivalently, though without the perspicuous operational semantics of an automaton, events may be modelled as the vocabulary of a regular expression, and complex events by the compositional (bracketed or sequential) substructures of regular expressions.

3.2 Regular models of rhythm

An obvious explicandum for regular models in the formal sense is rhythmic temporal patterning, which is also 'regular' in the informal sense. Rhythm is not only sequential but cyclic. First, rhythm is a sequential iteration of comparable events, interpreted with the temporal precedence relation. Second, hierarchies of rhythm patterns are interpreted in compositional cycles, as in the cyclical, i.e. recursive, interpretation functions of logical semantics and Generative Phonology (CHOMSKY & HALLE 1968). Such cycles have been a frequent topic of discussion in the application of prosodic models to prosody generation for speech synthesis (WAGNER 2002).

In both phonology and phonetics, three rhythmic patterns are generally distinguished, and associated with phonological constructs such as the syllable, the foot (stress or beat unit):

1. Unary rhythm: a sequence of regular beats, as in syllable timing and in sequences of the type in English *Slim Jill swam fast past Jim's boat*.
2. N-ary rhythm (with binary rhythm as the unmarked case) as two types of temporal sequence organised in a two-level hierarchy, the lower level consisting of a sequence (*foot*) of one strong and at least one weak (e.g. longer and shorter) element, and the higher level being a sequence of such feet:
    1. trochaic: initial element of the foot is strong, with the special case of a preceding *anacrusis* of weak elements (JASSEM, HILL & WITTEN 1984), as in *Little Johnny sang for supper*, based on morphological structure.
    2. iambic: final element of the foot is strong, as in *A fish can help to feed the cat*,

6 See: `people.umass.edu/partee/409/Is_Language_Infinite.pdf`

based on sentence structure.

These cases can be modelled by simple regular expressions:

Unary rhythm: b^+ i.e. *bb* *

N-ary rhythm:

trochaic: $(sw^+)^+$ i.e. $(sww^*)^+$

iambic: $(w^+s)^+$ i.e. $(ww^*s)^+$

In the trochaic and iambic cases the extent |w *| of the Kleene star '*' expression (i.e. the number of weak elements) is determined by typological (word and sentence stress, focus conventions), and performance (speech production) conditions. Generally, $1 < |w^*| < 4$ except in cases of extreme rhythm with additional tempo or time constraints (as in the metrical rhythm of a song). An example of typological conditions at word level is provided by highly inflecting languages with stressed stems and obligatory unstressed vocalic suffixes, which support a tendency to trochaic rhythm. Conversely, languages with inflectional prefixes or left positioned grammatical elements are expected to tend towards iambic rhythm. The iambic case has been demonstrated experimentally at sentence level in English (GIBBON 2003a, 2003b)

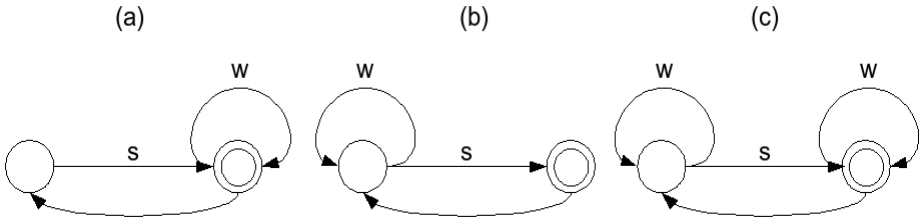

Figure 2: (a) trochaic, (b) iambic and (c) generalised rhythm automata.

. The regular expressions are visualised as finite transition networks in Figure 2(a) and (b). It is tempting to generalise all these types with a single regular expression, as visualised in Figure 2(c):

Universal rhythm: $(w^*sw^*)^+$

The expression describes arbitrary sequences of *w* and *s* which contain at least one *s*, but which evidently require strong further constraints provided by the typological categorial structures underlying trochaic (in English: morphological) and iambic (in English: syntactic) rhythmic tendencies.

The consequences of generalising the rhythm automata are potentially rather interesting:

1. The rhythm types reduce to typological (morphosyntax), pragmatic (focus) and performance constraints on the composition and therefore extent $|w^*|$ of the Kleene star weak sequences.
2. Trochaic and iambic types may combine by concatenation, with two main juncture types, as a blend *wsw* or a stress clash sequence *wssw*, and in other ways derived from these.
3. Anacrusis reduces to a special case of a combination of iambic and trochaic types.

In this approach, the unit *foot* turns out to be an epiphenomenon, perhaps simply an artefact, superimposed on sequences of strong and weak units as defined by the regular expressions

and constraints on the extent |*w**| of the Kleene star. On this analysis, there is apparently no need to introduce an interface based on the foot (minor rhythm unit, etc.), originally a metaphor derived from the speech of poetic composition and performance, as an additional constraint. Instead, constraints are imposed directly.

An example of a language for which analyses may benefit from such an approach is Polish, which defies attempts to classify it as foot or syllable timed: penultimate word stress supports the blend analysis, with prosodic word structures modelled as *w*sw*, i.e. without a Kleene star on the final weak element (GIBBON, BACHAN & DEMENKO 2007).

Morphosyntax is also behind the concept of anacrusis: examples cited in the literature are typically elements which are weak by default on grammatical grounds, such as grammatical words or prefixes, as in *take Grey to London* vs. *take Greater London* (JASSEM, HILL & WITTEN 1984), the former is prosodically parsed into temporal groups with anacrusis, as (*take)(Grey*)(*to London*), the latter without, as (*take*)(*Greater*)(*London*).

The approach to modelling temporal patterns and their relation to locutionary structures as regular languages thus has potential for relating prosody to typological features of languages such as 'right vs. left headedness', including prefixation and proclisis on the one hand, and suffixation and enclisis on the other.

In addition to its relevance for prosodic typology, another consequence of the present approach is the shift of focus to cognitive performance conditions for the development of a theory of emergent, cognitively constructed rather than empirically induced behavioural rhythm: perceived rhythm is modelled as a function of both top-down constraints and bottom up empirical temporal speech patterns (GIBBON 2006). However, this is a separate topic, and will not be addressed further here. Nor will the phonetic interpretation of the relational categories *s* and *w* be discussed further at this point, or the Absolute Time properties of these categories.

3.3 Regular models of intonation

There have been many regular models of intonation, for a numbr of langauges, but the most well-known is that of Pierrehumbert for English (PIERREHUMBERT 1980). It has been modified many times, and applied to many languages, but the basic properties remain the same. The regular expression version of this model is as follows (the non-alphabetic characters, including the asterisk, are plain terminal symbols and do not have any further formal meaning):

((%H | %L) ((H* | L* | L*+H | L+H* | H*+L | H+L*)$^+$ (H- | L-))$^+$ (H% | L%))$^+$

A visualisation of the regular expression as a finite transition network is shown in Figure 3. The elements of the vocabulary {%H, %L, H*, L*, L*+H, L+H*, H*+L, H+L*, H-, L-, H% , L%} are termed tones, and are shown with their syntagmatic functions in Table 2.

Each of the asterisks '*' in the finite automaton corresponds to a 'stress' position in conventional terminology, where the linguistic trigger for the stress is not specified - it could be word stress, sentence stress, or focus. More generally, the Pierrehumbert tones are a pair of stress and accent, in the Bolinger terminology, stress being an abstract structural position and accent being the pitch pattern occurring at this position.

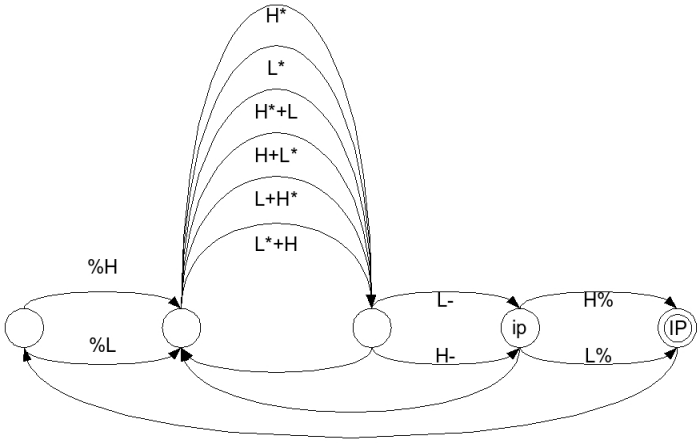

Figure 3: Finite transition network after PIERREHUMBERT *(1980).*

Table 2: Pierrehumbert 1980 tone inventory.

Function		*Form*
Initial boundary tones:		%H, %L
Accent tones:		H*, L*
		L*+H
		L+H*
		H*+L
		H+L*
Final boundary tones:	intermediate phrase (ip):	H-, L-
	intonation phrase (IP):	H%, L%

In terms of prosodic events, the Pierrehumbert tones may be formulated as pairs <*property*, *interval*>, more specifically as <*contour*, *location*> pairs, where the term 'location' refers to the interval component of events. The Pierrehumbert tones are pairs <*variable-accent*, *stress-location*> representing the phonetic interpretation of stress. In a pitch accent language, the pitch pattern would be constant, but would occur in different locations: <*fixed-accent*, *accent-location*>, while a tone language would not be associated with stress but with a tone-bearing unit (TBU), and the pitch contour would be variable <*variable-tone*, *tone-bearing-unit*> (cf. HYMAN 2007a for a critical discussion of stress, tone and accent typology problems). These tonal pairs define a multidimensional typological space in which location types and the shapes and inventories of pitch patterns are located.

Given the interpretation of the Pierrehumbert tones as <*contour*, *location*> pairs, interpreted for English as Bolingerian <*accent, stress*> pairs (BOLINGER 1985), the topology of the main iteration in the automaton turns out to correspond to the 'rhythm group', 'major rhythm group', or 'body' of an intonation group in traditional discussions of intonation. On the other hand, the pair of Intermediate Phrase and Intonation phrase boundary markers, together with the final <*accent*, *stress*> pair, turns out to be a component of the 'nuclear stress' and 'nuclear tone' contours of traditional intonation descriptions.

It is necessary to provide a perspicuous meta-semantics for the regular expression for the purpose of assignment to locutionary units by the media-semantic interpretation function. For this purpose it is useful to define a syntagmatic grouping of the constituents, modelled by a tree. The topology of this network is defined by the bracketings of the regular expression, which in turn are determined by the iterations required, and modelled as a tree of finite depth (Figure 4).The hierarchical structure has a mix of finite depth and unidirectionally branching tone sequence substructures, and can be represented either by a context-free (Type 2) or a regular (Type 3) grammar. A Type 2 grammar is convenient for the semantics of assignment to locutionary items (morphemes, words, phrases, etc.) but not strictly necessary; the relevant formal language is Type 3.

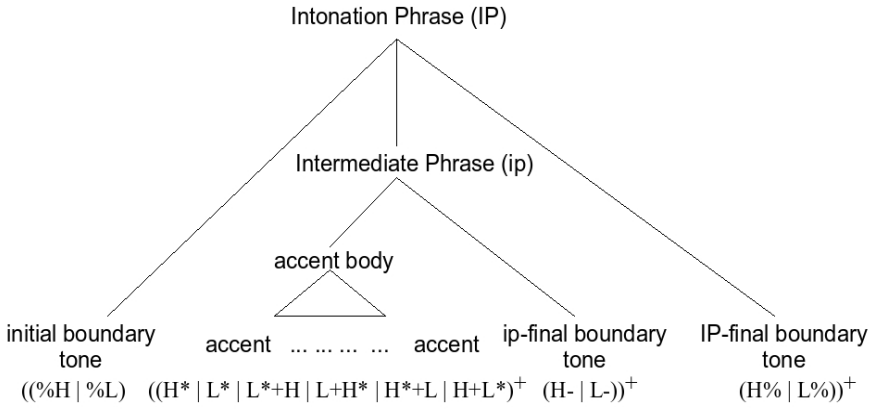

Figure 4: Hierarchical topology of Pierrehumbert finite state automaton.

Cognosci of the intonation literature will recognise the hierarchical topology at this level of abstraction as being quite similar to that of many earlier approaches to the modelling of intonation, such as those analysed by GIBBON (1976). There were also earlier analyses in finite state terms, such as those of the Dutch school in the 1970s (reported in 'T HART, COLLIER & COHEN 1990) and in work since the 1960s by FUJISAKI, but the Pierrehumbert model is the one which is most easily related to both phonological and phonetic categories. Curiously, none of the earlier finite hierarchy based approaches (cf. the approaches represented in CUTLER & LADD 1984) attempted finite state modelling, though REICH (1969) had already suggested this structure for intonation patterns.

As with rhythm automata, there are further constraints to be defined for the tonal sequences, based on grammatical and pragmatic considerations. Gibbon showed that a meaningful constraint on regular accent sequence patterns may occur with a rhetorical interpretation (GIBBON 1984): the constraint is to limit all accents in a sequence to the same type, for example rising [L*+H] or falling [H*+L]. A constraint of this kind is shown in the following example, where the grave accent '`' represents [L*+H] rising and the acute accent '´' represents [H*+L] falling accents, a single '\' represents a [L-] ip-boundary tone, and '\\' represents a sequence of ip-final [L-] and IP-final [L%] boundary tones:[7]

> Lìttle hèdgehog trùndled alòng\, thròugh the lèaves and the grèen stuff ìn the wòod\, lòoking for sòmething nìce to èat\\. He'd néver béen outsíde of the wóod before\\.

This accentuation pattern can be described by the regular expression

7 Overheard in a children's story related by 'Uncle Bill' on BFBS radio, Germany, sometime in the 1980s.

$$r^{+}\, b\, f^{+}\, b$$

where *r* and *f* stand for rising and falling accents, respectively, and *b* stands for boundary tone subseques demarcating Intermediate Phrases or Intonation Phrases. Such *falling*$^{+}$ or *rising*$^{+}$ *falling* patterns are frequently to be found within a single Intonation Phrase, with rising sequences typically as topic-introducing stretches and falling tone sequences as topic-elaboration stretches, as in the 'little hedgehog' example.

3.4 Regular models of tone: terracing in Niger-Congo languages

As with models of rhythm and intonation, it has been shown that tonal patterns in Niger-Congo and Sino-Tibetan languages can be modelled as regular languages, generalised as finite state automata and visualised as finite state networks. The starting point for these modelling conventions is to be found in the application of Metrical Phonology to tonal patterning.

A number of Niger-Congo languages, such as the closely related Kwa languages Baule and Anyi, have high and low level tones, *H L*, whose height is interpreted phonetically as relative to the height of preceding or following tones, and as being sometimes dependent on segmental categories such as tone-blocking and tone-lowering consonants. A common phonetic interpretation function in tone languages is *tone-terracing*, in which a drop in pitch from a H^{+} sequence to a L^{+} sequence is larger than the increase in pitch from a L^{+} sequence to a H^{+} sequence. This asymmetry leads to a stepwise sequence of lowered 'tone terraces' (Figure 5).

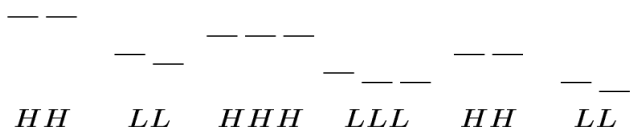

Figure 5: Idealised visualisation of a terraced tone sequence consisting of 3 terraces, each containing a H demi-terrace and a L demi-terrace.

In the Metrical Phonology literature (CLEMENTS 1981), sequences of this kind are mapped to right-branching trees of the kind shown in Figure 6. Metrical Phonology descriptions emerged at a time when the focus in linguistic descriptions was turning to representations and constraints on representations, rather than on rules. Nevertheless, it is instructive to enquire about appropriate formal grammars for generating or accepting such representations, since this provides essential information about the complexity of the representations, a line of enquiry which was no longer pursued in linguistics at that time.

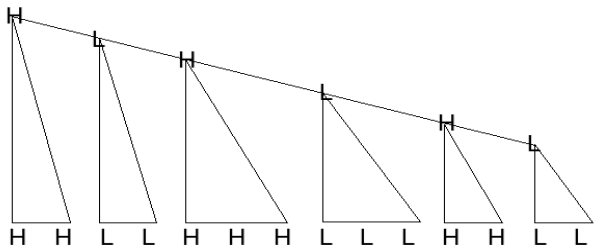

Figure 6: Right-branching tone-terrace tree representation (internal right-branching structure of demi-terraces represented by triangles).

Oddly, before the background of Generative Grammar, no grammars were offered

which would generate these structures. However, examination of the metrical tree representations shows that the trees are right branching, with finite depth subcomponents. Neither right-branching trees nor trees of finite depth require more complex devices than regular (Type 3) grammars and finite automata. Gibbon (Gibbon 1987, 2001) introduced automata of this type which generalised over right-branching metrical trees, yielding branching-neutral representations (Figure 7).

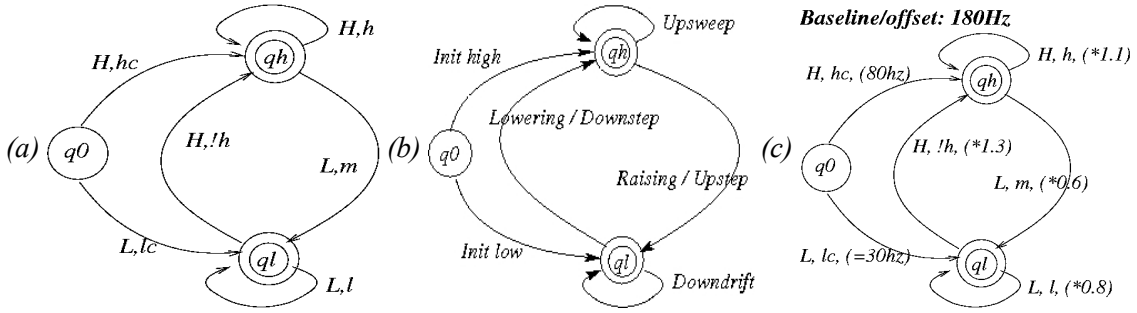

Figure 7: (a) Basic generalised finite state transducer model for phonetic interpretation of tone in two-tone Niger-Congo languages, (b) mapping of the topology of the model on to traditional syntagmatic tone sandhi terminology, (c) one possible interpretation of the transition functions in terms of absolute initial frequencies and an asymptotic declination function.

The tone automaton in its simplest form, for two-tone languages, has a start state which branches into H and L tone paths, and two further states, a state at which H tones terminate and a state at which Low tones terminate. Each of these two states permits iteration to itself (which defines a demi-terrace), and a further iteration cycle is defined from each of the states to the other and back again (which defines a full terrace). These two kinds of iteration correspond to the two levels in the metrical hierarchy shown in Figure 6. On the transition from the L state to the H state, the phonetic H is downstepped; other phonetic functions such as tonal assimilations can be associated with the other transitions, and differ from language to language.

The tree model for the topology of these tonal automata differs fundamentally from the finite-depth centre-embedding-plus-preterminal-right-branching[8] tree model for the topology of the Pierrehumbert-type intonational patterns. The topology of the two trees, and of the automata networks, clearly illustrates a different kind of two-level hierarchy for the tone sequences, which shows a disjunction of iterations which is quite unlike the intonational patterns. The disjunction is explicit in the corresponding regular expression, which is for this reason more inelegant than other representations, in this case, and does not immediately suggest the oscillatory character of the model which is evident in the visualisation as a finite state network. However, the right-branching structure of the regular expression does express something of the topology of the automaton, and of the structure of the metrical tree. The regular expression is a disjunction $A|B$ of two very similar regular expressions, one starting with the 'InitHigh' category and one with the 'InitLow' category (the high categories are represented here by H, the low categories by L), and sharing sub-expressions which are inverses of each other:

8 Left-branching would also be an option: for any right-branching regular grammar there is a left-branching regular grammar which generates the same language, and vice versa. The automaton notation is neutral with respect to left or right branching.

A | *B*

A=inithigh (upsweep | upstep downdrift* downstep)* (upstep downdrift*)$^{?}$

B=initlow (downdrift | downstep upsweep* upstep)* (downstep upsweep*)$^{?}$

The asterisk denotes the Kleene star (zero or more iterations), and the question mark denotes zero or one iterations. The correspondence between the regular expression and the automaton was verified using an automatic visualiser as shown in Figure 8.[9] The coding used in the visualiser is shown in

Table 3: Terminal symbol encoding for visualisation automaton 'reAnimate'.

inithigh	*A*	*initlow*	B
upsweep	*U*	*downdrift*	D
upstep	*L*	*upstep*	H

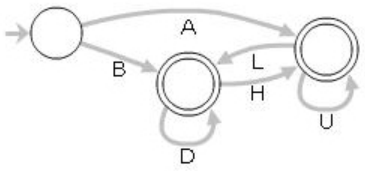

Figure 8: Automatic visualisation of the regular expression corresponding to the tone automata, coded as:

A (U | L D* H)* (L D*)$^{?}$ | B (D | H U* L)* (H U*)$^{?}$

The two disjuncts in the regular expression have the same structure, differing only in terminal elements. Each disjunct has a clear structure, illustrated here with the *A* disjunct:

1. The initial transition, 'inithigh'.
2. A main iterating section, '(upsweep | upstep downdrift* downstep)*', permitting iteration of either 'upsweep' alone, or the main loop 'sequence upstep downdrift*' 'downstep' (including iteration of 'downdrift'), or any combination of these two disjuncts,
3. A final section, '(upstep downdrift*)$^{?}$', permitting one or more occurrences of the 'upstep downdrift*' sequence, enabling termination with 'upstep' and zero or more occurrences of 'downdrift'.

The *B* section has the opposite specification, in which 'upsweep' is replaced with 'downdrift', 'downdrift' with 'upsweep', 'upstep' with 'downstep' and 'downstep' with 'upstep'. An equivalent regular grammar is given in Table 4.

9 In the regular expression as originally printed (Gibbon 2001), the *A* disjunct, which can be recoded here as 'A(U(LD*H)*)*(LD*)', had a typo: missing '*' after 'U' (note that in this notation the final parentheses have the same function of optionality as '?' in the present notation). The U* is actually redundant; the present formulation with '|' is better. The starred parenthesis around 'LD*H' is also redundant and can be removed. Thanks to Christoph Schillo and Ben Hell for discussion of optimisation of the regular expression, and to Jolanta Bachan for discussion of optimisation and visualisation of the regular expression with Oliver Steel's Regular Expression visualiser *reAnimator*: <http://osteele.com/tools/reanimator/> (consulted 2008-08-08).

Table 4: Regular (Type 3) grammar for two-tone terracing automata. In this notation, 'S' is the initial symbol, 'H' and 'L' are non-terminal symbols, aand the lower case items are terminal symbols.

S	→	inithigh H
S	→	initlow L
H	→	upsweep H
H	→	upstep L
H	→	upsweep
L	→	downdrift L
L	→	downstep H
L	→	downdrift

Gibbon discusses specific details (GIBBON 2001), in which more complex Niger-Congo tone systems are also described. An interesting implementation option for such grammars is offered by the default inheritance language DATR (Table 5), where 'lc' and 'hc' stand for 'low constant' and 'high constant', respectively, and 'dh' stands for 'downstep high'.

Table 5: Default inheritance grammar for two-tone terracing automata.

```
Tone:
        <l> == lc Tone_low:<>
        <h> == hc Tone_high:<> .
Tone_high:
        <h> == h <>
        <l> == l Tone_low:<>
        <> == .
Tone_low:
        <l> == l <>
        <h> == dh Tone_high:<>
        <> == .
```

The syntax of this restricted use of the DATR language is straightforward: the representation can be interpreted as a finite state automaton, with transitions from the same node grouped together under this node, and with the '==' inheritance symbol as the transition function to another (or cyclically to the same) node. Alternatively, the notation can be interpreted as a left-branching regular grammar, in which rules with the same left-hand side non-terminal symbol are grouped under this symbol, the terminal symbol is to the left of the '==', and the target non-terminal symbol is to the right of the '=='. The empty cases denote empty input and output, i.e. the end of the string.

DATR has symbolic processing only, but a modification of this implementation has also been used to generate textual numerical phonetic output, which is then interpreted by a conventional language or a spreadsheet in order to produce f0 plots.

3.5 Regular models of tone: Mandarin

The most well-represented tone language in the literature is Mandarin Chinese. Jansche developed a model of the phonetic interpretation of tones in the Tianjin variety of Mandarin (JANSCHE 1998; cf. Table 6). There are four *phonemic* tones. The *phonetic* tones are represented by numbers, 1 being the lowest and 5 the highest (the phonetically bracketed '[21]' signifies that only a surface 21 tone can trigger this allotone, not an underlying lexical 21.

Table 6: Tianjin Mandarin tone inventory.

	Phonemic tone	*Phonetic tone in isolation*	*Allotone*
1.	A1	21	→ 213 / __ 21
2.	A2	45	→ 45
3.	B	213	→ 45 / __{B, 213}
4.	C	53	→ 45 / __ [21], 21 / __ 53

Jansche demonstrated that the complex tone combinations (tone sandhi), in which the phonetic interpretations of tones are partly conditioned by preceding tonemes, partly by specific allophones of preceding tonemes, can also be modelled as a regular language, along similar lines to the regular languages already described for Niger-Congo languages. A partial grammar for this regular language is modelled by an automaton accounting for the syntagmatic restrictions on the first three tones, visualised by a finite state network. The restrictions on the fourth tone are not modelled (Figure 9).

The topology of the Tianjin network is not only different from the automaton which models the two-tone terraced Niger-Congo languages, but, like these, is also very different from the intonation networks. The Tianjin network shares one or two features with the terraced tone network, including the assignment of exactly one node to each tone, in addition to a start node, and the connection of all nodes with each other, with the one exception that there are far more nodes, nodes 2 and 3 are not connected. None of the connections form the kind of two (or more) state oscillatory cycle as in the Niger-Congo model.

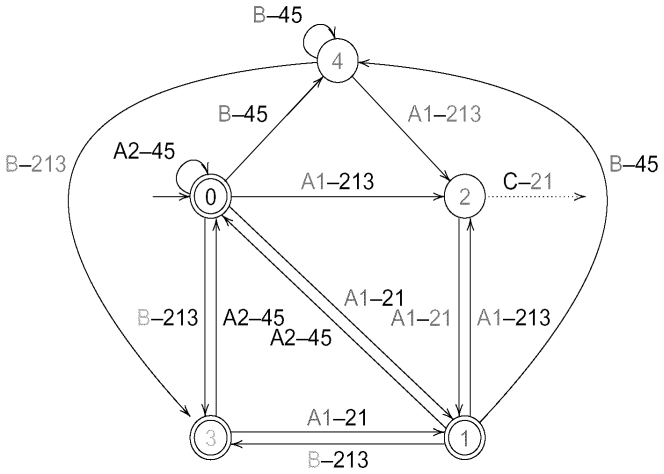

Figure 9: Partial finite state network for Tianjin Mandarin tone (JANSCHE 1998).

4 Prosodic Rank Theory

4.1 Ranks

The starting point of this paper was *Dual Interpretation Theory* as a semiotic framework, and at this point the argument returns to the semiotic framework in order to provide a context for the patterns modelled in the preceding sections. Before the advent of Chomskyan linguistics, in which the focus of attention was narrowed down under the influence of formal logical

concepts of the syntax of theories and their interpretation, primarily to sentences, words and their constituents, a broader conception of linguistics was more common, and is being revived, partly under the influence of formal dialogue modelling in speech technology.

Linguists such as Pike, Jespersen and, somewhat after the inception of the Chomskyan paradigm, Halliday, contextualised their structural descriptions within a larger functional framework of a hierarchy of types of unit, for which Jespersen's term 'rank' (JESPERSEN 1924) has become widely used. This approach will be used here to systematise the compositional role of prosody in *Dual Interpretation Theory*.

4.2 Compositionality in Prosodic Rank Theory

The rank hierarchy which is particularly useful for modelling prosody consists of the categories *phoneme – morpheme – simplex word – derived word - compound word – phrase – clause – sentence – text/turn – dialogue*. Categories at each of these ranks have their own compositional principles. Each rank has the two kinds of 'semantic' interpretation described in *Dual Interpretation Theory*. The obvious semantic interpretation is the familiar *object semantic* interpretation – the conventional model semantics of logic or the logical form of linguistics. The less obvious interpretation (to all except phonologists and phoneticians) is the *media-semantic* interpretation in terms of the physical communicative events of speech, gesture and writing.

To review the framework: the *phonetic interpretation* of generative phonologies is a special case, but it also covers the *visual interpretation* of writing and visible gesture. The media-semantic interpretation function is, like the object-semantic interpretation function, also a semantic interpretation function in the strict sense of model semantics: it maps a language (in conventional linguistic terms: the underlying representation) to a specified domain, in this case the domain of the media world, which includes the phonetic domain.

The object-semantic and media-semantic interpretation functions apply to each level in the rank hierarchy, yielding the overall compositional and interpretative *Prosodic Rank Theory* which is visualised in Figure 10 together with dual interpretation models at each rank. The rank hierarchy is simplified in the figure for presentation purposes. In this comprehensive form, the approach is referred to as *Ranked Dual Interpretation Theory* (*RDIT*).

The concept of rank adds granularity to the core ontology introduced in the initial sections of the present study. Signs are compositional (with modifications in the case of lexicalised complex items such as idioms), and their object-semantic and media-semantic interpretations reflect this compositionality: the meaning of a sentence is a function of the meanings of its parts, and the pronunciation or spelling of a sentence is a function of the pronunciations or spellings of its parts. The exceptions which prove the rule are idioms, where the holistic contribution of the whole item to the meaning overrides the compositional meanings of the parts, and the lexicalisation processes which render words like 'beside' (by side) and 'husband' (hus bonda) phonologically and orthographically opaque.

Further, each rank has its own specific compositional principles. Thus, the composition of words from morphemes and morphemes from phonemes (or other appropriate segmental units) has a different grammar from the composition of phrases from words, or sentences from words, texts or turns from sentences, and dialogues from texts or turns. Consequently, rank-specific category types are defined, and – in terms of formal grammars – the terminal symbols at any rank are the starting symbols of the rank immediately below. There are exceptions: downward type conversions, traditionally known as 'rank shift', such as 'do-it-yourself shop', in which the imperative sentence 'Do it yourself!' functions as an adjective.

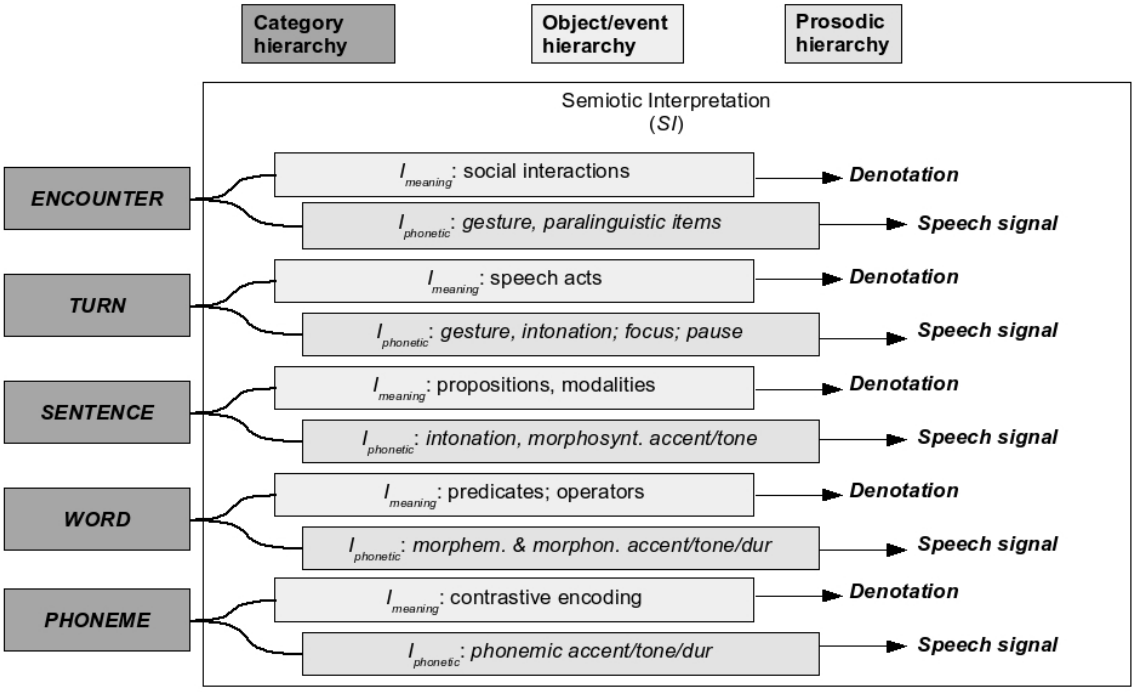

Figure 10: Rank hierarchy composition and interpretation: the 'signifiant-signifié' relation explicated as a hierarchy of function dyads.

The *RDIT* framework has a number of advantages over other approaches. First, it is compatible not only with structural concepts but also with functional concepts in linguistics. Second, it provides an account of language, and specifically of prosody, which is coherent from phonemic through to discourse units. Third, this coherence is inherently superior to the traditional linear *syntax-semantics-pragmatics* model, which links non-comparable categories, and in which it is unclear where to fit in areas like prosody and idiomaticity. In the *RDIT* framework, prosody is generalised phonetic interpretation at every rank, and idiomaticity is lexicalisation of complex structures at every rank. A special application of the *RDIT* framework is *Prosodic Rank Theory*, which will be summarised in the following section.

4.3 Interpretativity in Prosodic Rank Theory

The prosodic ranks are the 'media semantic' interpretations of the compositional categories at each rank. The 'object semantic' and 'media semantic' interpretation functions at each rank level follow the compositionality principle:

1. The object semantic interpretation of a phoneme is simply to be a contrastive encoding for morphemes, and its phonetic media semantic interpretation is in terms of classic phonetic features defined by the International Phonetic Alphabet (or other appropriate alphabet).
2. The default object semantic interpretation of a word is a predicate (for lexical words) or an operator (grammatical words), and the phonetic media semantic interpretation of a word is a compositional function of the phonetic media semantic interpretation of its morphological parts, and these in turn are a

compositional function of the phonetic media semantic interpretation of the lower ranking phonemic parts. The compositional phonetic media semantic interpretation function for words includes prosodic components: the assignment of word stress and other lexical and 'post-lexical' operations on the lower rank phonetic interpretation.

3. The default object semantic interpretation of a sentence is a proposition, together with a modal (epistemic, doxastic or deontic) operator and a mood (illocutionary) operator. The phonetic media semantic interpretation of a sentence is a compositional function of the phonetic media semantic interpretation of its parts, and includes the assignment of an intonation pattern. Similarly, the spelling of a sentence is a compositional function of the spelling of the parts, whereby the function includes the setting of sentence-initial capitalisation and punctuation.

Mutatis mutandis the same interpretativity principle applies to the other ranks: in *Prosodic Rank Theory*, prosodies such as tone, pitch accent and word stress may operate phonemically or morphologically (in derivation, compounding and in morphosyntax, depending on the language concerned), and prosody is therefore subsumed under a generalised concept of phonetic interpretation as a rank-structured denotational semantic model.

The ranks are an ordering in terms of the default sizes (temporal extent) of communicative units of speech, and therefore of the temporal extent of these communicative units. Time, in terms of temporal extent, is thus a semiotic category of media interpretation and semantic interpretation at all rank levels: time is an object-semantic category according to which events referred to in speech and text are organised, and time is a media-semantic category according to which speech, gesture and text production, transmission and reception are organised and contextualised. The media-semantic notion of time at each different rank is the specific defining characteristic of the study of prosody: the temporal organisation of segmental and melodic patterns.

5 Outlook

It has been shown that a wide range of prosodic patterns can be modelled as regular languages with grammars which are operationalised as finite state transducers. These patterns include rhythm patterns, intonation patterns, and tone sequences of Niger-Congo and Tianjin Mandarin Chinese as a representative of Sino-Tibetan.

Models of this kind have well-understood declarative properties in terms of Event Logic, well-understood procedural properties in terms of automata theory, and a well-understood operational semantics in terms of computational implementations. Further, finite state devices implement well-understood modelling conventions for temporal precedence relations in speech. These properties make such devices both linguistically interesting and technologically useful.

But the place of regular languages (and the generalisation of these as regular relations (KAPLAN & KAY 1994) in the overall scheme of language has not always been clear. Chomsky (CHOMSKY 1957) famously claimed that finite state devices are not adequate for modelling natural languages: 'English is not a finite-state language'. Whether this is true or not, the claim has certainly to be relativised for specific ranks: finite state devices are now known to

be adequate for modelling at the phonological and morphological ranks, as has been amply shown in the computational linguistics literature on finite state syllable structure (e.g. CARSON-BERNDSEN 1998) and in Finite State Morphology (e.g. BEESLEY & KARTTUNEN 2003). Simple sentences can also be modelled by such devices (including disjoint constituents in simple sentences), though complex sentences with centre-embedding (e.g. subject relative clauses in English) and indexed constructions (as with 'respectively' conjunction pairs in English) evade conventional finite state modelling unless an arbitrary finite depth of embedding is imposed.

The roles of these formal devices differ considerably on the compositional rank scale of lexical, morphosyntactic, syntactic, textual and discoursal construction, but finite state devices offer a solid perspective for a tractable integration of prosodic patterns and processes at these prosodic ranks: Mandarin has phonemic functions, for which an assignment of tonal transitions to syllables in an extension of the regular relation concept is sufficient (the role of stress in tone languages such as Mandarin has not been considered here). Niger-Congo languages have phonemic tone and also morphosyntactic tone which marks inflections and complementiser constructions, as well as nominal compoounding of the kind known in African linguistics as associative constructions. There are indications that at least some Sino-Tibetan, in particular Tibeto-Burman, languages have similar properties (EVANS 2002; HYMAN 2007b). And all languages apparently have intonational forms and functions (HIRST & DI CRISTO 1998) which can be assigned to different levels in *Prosodic Rank Theory*. It would go well beyond the scope of this contribution to go into any more detail on these points, and a sketch of the framework within which the formalisation of further details can take place will have to suffice.

In summary: the major result of this study is a characterisation of prosodic typology which goes beyond previous studies (cf. contributions in GUT & GIBBON 2001) in terms of events and the compositional and interpretative *Prosodic Rank Theory*, in the context of *Ranked Dual Interpretation Theory,* which introduces both object-semantic and media-semantic interpretations in the sense of formal model theory. The model presented here provides a principled and semiotically well-founded basis for integrating prosody into the rest of the linguistic world, and contrasts starkly with the traditional approaches to prosody modelling, which present a linguistic archipelago of unrelated lexical, syntactic, semantic and pragmatic functions, from phonemic tone through accent and focus to discourse intonation.

6 References

BARWISE, JON & JOHN PERRY. (1983) *Situations and Attitudes*. Cambridge: MIT Press.

BATÓG, TADEUSZ (1967). The Axiomatic Method in Phonology. London: Routledge & Kegan Paul.

BEESLEY, KENNETH R. & LAURI KARTTUNEN (2003). *Finite State Morphology*. Stanford: CSLI Publications.

VAN BENTHEM, JOHAN (1988[2]). *A manual of Intensional Logic.* Stanford: CSLI Publications.

BOLINGER, DWIGHT L. (1986). *Intonation and its Parts: Melody in Spoken English*. London: Edward Arnold.

BIRD, STEVEN & EWAN KLEIN (1990). "Phonological Events." *Journal of Linguistics* 26, pp 33-56.

BROWMAN, CATHERINE P. & GOLDSTEIN, LOUIS (1992). "Articulatory phonology: an overview." *Phonetica* 49 (3-4), pp 155-180.

CARSON-BERNDSEN, JULIE (1998). *Time Map Phonology. Finite State Models and Event Logics in Speech Recognition.* Dordrecht: Kluwer Academic Publishers.
Carson-Berndsen, Julie & Dafydd Gibbon (1992). "Event relations at the phonetics/phonology interface." COLING 1992, pp 1269-1273.
Chomsky, Noam (1957). *Syntactic Structures*. The Hague: Mouton.
CHOMSKY, NOAM & MORRIS HALLE (1968). *The Sound Pattern of English.* New York: Harper & Row.
CLEMENTS, G. N. (1981). "The hierarchical representation of tone features." Harvard Studies in Phonology 2. Distributed by IULC.
Connell, Bruce (2001). "Downstep, downdrift and declination." In: GUT & GIBBON (2001).
CUTLER, ANNE & D. ROBERT LADD, eds. (1983): *Prosody: Models and Measurements*. Heidelberg: Springer-Verlag, pp 123-140.
DZIUBALSKA-KOŁACZYK, KATARZYNA (2002). *Beats-and-Binding Phonology.* Frankfurt am Main: Peter Lang.
EVANS, JONATHAN (2002). "'African' tone in the Sinosphere." Report, Institute of Linguistics, Academia Sinica
GIBBON, DAFYDD (1976). *Perspectives of Intonation Analysis*. Bern, Lang.
GIBBON, DAFYDD (1983). "Intonation in context: an essay on metalocutionary deixis." In: Gisa Rauh, ed. *Essays on Deixis*. Tübingen: Narr.
GIBBON, DAFYDD (1987). "Finite state processing of tone languages." In: *Proceedings of European ACL*, Copenhagen, pp 291 - 297.
GIBBON, DAFYDD (1992). "Prosody, time types and linguistic design factors in spoken language system architectures." In: G. Görz, ed., *KONVENS '92*. Berlin, Springer, S. 90-99.
GIBBON, DAFYDD (1994). "Intonation as an adaptive process". In: Dafydd Gibbon & Helmut Richter, eds., *Intonation, Accent and Rhythm. Studies in Discourse Phonology.* Berlin: Walter de Gruyter, pp 165-192.
GIBBON, DAFYDD (2001): "Finite state prosodic analysis of African corpus resources". In *EUROSPEECH-2001*, pp 83-86.
GIBBON, DAFYDD (2003a). "Computational modelling of rhythm as alternation, iteration and hierarchy". *Proceedings of the International Congress of Phonetic Sciences*, Barcelona, August 2003, III: 2489-2492.
GIBBON, DAFYDD (2003b). "Corpus-based syntax-prosody tree matching". In: *Proceedings of Eurospeech 2003*, Geneva.
GIBBON, DAFYDD (2006). "Time Types and Time Trees: Prosodic Mining and Alignment of Temporally Annotated Data." In: Sudhoff, Stefan & al. (2006). *Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter., pp. 281-209.
GIBBON, DAFYDD, JOLANTA BACHAN & GRAZYNA DEMENKO (2007). "Syllable timing patterns in Polish: results from annotation mining." *Proceedings of Interspeech/Eurospeech 2007*, Antwerp.
GOLDSMITH, JOHN 1990. *Autosegmental and metrical phonology*. Oxford: Basil Blackwell.
GUT, U. & DAFYDD GIBBON, eds. (2002). *Typology of African Prosodic Systems.* Bielefeld: Bielefeld Occasional Papers inTypology 1.
'T HART, JOHAN, RENÉ COLLIER & ANTONIE COHEN (1990). *A Perceptual Study of Intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.
HIRST, DANIEL & ALBERT DI CRISTO (1998). *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.

HYMAN, LARRY M. (2007a). "How (not) to do Phonological Typology: The Case of Pitch-Accent." UC Berkeley Phonology Lab Annual Report 2007.
HYMAN, LARRY M. (2007b). "Kuki-Thaadow: An African Tone System in Southeast Asia ." UC Berkeley Phonology Lab Annual Report 2007.
JANSCHE, MARTIN (1998). "A Two-level Take on Tianjin Tone." In: Gert–Jan Kruijff & Ivana Kruijff–Korbayov, eds., *Proceedings of the Third ESSLLI Student Session*, 10th European Summer School on Logic, Language and Saarbrücken, Germany, pp 162–174.
JAKOBSON, ROMAN O. (1960). "Closing Statements: Linguistics and Poetics." Thomas A. Sebeok, ed. *Style In Language*. Cambridge Massachusetts: MIT Press.
JASSEM, WIKTOR, DAVID R. HILL & IAN H. WITTEN (1984). "Isochrony in English Speech: its Statistical Validity and Linguistic Relevance." In: Dafydd Gibbon & Helmut Richter, eds., *Intonation, Accent and Rhythm. Studies in Discourse Phonology.* Berlin: Walter de Gruyter, pp 203-225.
JESPERSEN, OTTO (1924). *The Philosophy of Grammar*. London: George Allen & Unwin.
KAPLAN, RONALD & MARTIN KAY (1994). "Regular models of phonological rule systems." In *Computational Linguistics* 20, pp. 331-378.
KAYE, JONATHAN (1989). *Phonology: A Cognitive View*. Hillsdale NJ: Lawrence Erlbaum Associates.
LADD, D. ROBERT (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
MITKOV, RUSLAN, ed. (2003). *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press.
PIERREHUMBERT, JANET (1980). *The phonology and phonetics of English intonation.* PhD thesis, MIT. Distributed 1988, Indiana University Linguistics Club.
POPPER, KARL (1965). "Of Clouds and Clocks. An approach to the problem of rationality and the freedom of man." Lecture, published in Karl Popper (1972), *Objective Knowledge*. Oxford: The Clarendon Press.
REICH, PETER A. (1969). "The finiteness of natural language." *Language* 45, pp 831-43.
STRAWSON, PETER F. (1959). *Individuals. An Essay in Descriptive Metaphysics*. London: Routledge.
TILLMANN, HANS G. (1980), *Phonetik. Lautsprachliche Zeichen, Sprachsignale und lautsprachlicher Kommunikationsprozeß*. Stuttgart: Klett-Cotta.
WAGNER, PETRA (1998). "Mutual Constraints at the Phonetics-Phonology-Interface." *Proceedings of the 4th Conference on Natural Language Processing KONVENS-98*. *Computer Studies in Language and Speech* Vol. 1: 1998,S. 207-212.
WAGNER, PETRA (2002). *Vorhersage und Wahrnehmung deutscher Betonungsmuster*. Dissertation, U Bonn.