Prosodic information in an Integrated Lexicon

Dafydd Gibbon

Universität Bielefeld, Germany

gibbon@spectrum.uni-bielefeld.de

Abstract

A Integrated Lexicon framework is proposed for combining insights from current separate prosodic research paradigms (language resource creation, experimental psycholinguistics, rules for speech synthesis, sentence search space restriction in speech recognition, prosody in discourse, emotional prosody) in the medium term into work in general and computational linguistics. Three principles for integrating autonomous prosodic information into the lexicon are proposed (the Compositionality Principle, the Semiotic Principle, and the Metalocutionary Principle), and used as a framework for developing hierarcical attribute-value (AV) lexical representations for nuclear contours and the prosody of discourse particles.

1. Lexical prosody for intonation

The representation of the basic inventory of prosodic units and their combinatorial, semantic and phonetic properties is an open question, despite much attention over many decades. For present purposes, this inventory of prosodic units is termed a *prosodic lexicon*, and a lexicon of words (or idioms) a *locutionary lexicon*. A proposal is presented for modelling autonomous lexical prosodic items and relating them to locutionary items in an integrated attribute-value (AV) based lexicon. The proposal is related to work by Aubergé on multiparametric prototypes and Steedman on lexical functions in Combinatory Categorial Grammar. Current paradigms of prosodic description concentrate heavily on other issues such as the extensive prosodic annotation and experimental modelling of prosodic patterns in speech corpora, procedural rule formulation for speech synthesis, prosody for restricting sentence search space in speech recognition, prosody in discourse, and prosody in emotion modelling. Each approach continually uncovers new insights, but it is unclear how they relate to each other.

Consequently the present contribution addresses the issue of integrating prosodic information into lexicalist approaches to language representation and computation. The approach is still fragmentary, but the medium-term goal is develop a formal representation for lexical prosody as a basis for compositional descriptions of more complex prosodic patterns, and as a reference point for combining results of different fields.

To provide a non-*ad hoc* lexical representation involves addressing three basic principles which I identify as follows: the *Compositionality Principle* (CP), the *Semiotic Principle* (SP), the *Metalocutionary Principle* (MP). The domain of the present contribution is lexical prosody for intonation, rather than lexical tone or lexical accent placement.

1.1. The Compositionality Principle (CP)

CP: *Prosodic patterns are grounded in a prosodic lexicon and are projected compositionally from prosodic lexical items and their combinatorial properties to larger prosodic patterns.*

The prosodic CP is characterised as follows for a compositional prosodic item $P = construction_{pros}(p_1, \dots p_n)$: $propPros_i(P) = composition(propPros_i(p_1), \dots propPros_i(p_n))$

The oldest explicit compositional approach to prosody is the *stress cycle* of generative phonology, in which "stress patterns" are projected from the lexicon by the Compound Stress Rule (CSR), and from syntactic structure by the Nuclear Stress Rule (NSR) as a numerical coding. The empirical basis of this coding has repeatedly been called into question.

In the past decade, Steedman [16] has paid most attention to the issue of projecting compositional prosody from the lexicon. Steedman proposes a set of accent and boundary tones with specific structural meanings and combinatorial properties as functions which map boundaries into utterances, yielding (with the syntax of Combinatory Categorial Grammar) hierarchical prosodic patterns corresponding to a hierarchy of foreground and background utterance constituents:

Accent tones:	L+H*	:= Theme/Bh
	H*	:= (Utterance/Theme)/bl
	H*	:= (Utterance\Theme)/Bl
Boundary tones:	LH%	:= Bh
	LL%	:= Bl
	L%	:= bl
Boundaries:	bl	intermediate phrase boundary
	Bl	intonation phrase boundary

Steedman's approach, and and the tone-level lexical representation, is embedded in a long history of prosodic phonologies, relating particularly to work by Selkirk [15], Pierrehumbert & Hirshberg [13] and Bird [2]. The approach accounts not only for syntactic constraints on prosodic patterning but also for divergences between prosody and syntax, and has been successfully implemented in prosody generators for speech synthesis. The selection of prosodic lexical categories discussed by Steedman is limited, perhaps because of the formal speech style of speech synthesis.

Another approach to compositionality is Finite State Prosody, represented in work by Fujisaki, 't Hart & Cohen, Pierrehumbert, Gibbon in the 1970s and 1980s, and more recently in work by Ladd [12], Jansche [10], and Gibbon [6]. In Finite State Prosody, a basic vocabulary of tones is mapped by a Finite State Automaton (FSA) into a *regular set* of tonal sequences. The tonal sequences may be mapped into allotone sequences by means of a Finite State Transducer (FST), the set of tone-allotone pairs and sequence pairs constituting a *regular relation* [11].

1.2. The Semiotic Principle (SP)

SP: *Prosodic patterns have a semiotic dual interpretation consisting of a compositional mapping to a semantic/pragmatic domain and a compositional mapping to a phonetic domain.*

The SP is characterised as follows, for $P = construction_{pros}(p_1, ...p_n)$:
$< propPros_{sem}(P), propPros_{phon}(P) >= composition(< propPros_{sem}(p_1),\ propPros_{phon}(p_1) >, \ldots\ < propPros_{sem}(p_n),\ propPros_{phon}(p_n) >$

In other words, a dual semiotic property of a complex unit is a function of the dual semiotic properties of its parts.

Steedman's analysis is a clear example of semantic interpretation side according to this principle, and his incorporation of a mapping to Taylor's *Tilt Model* of pitch patterning in recent work also introduces a dimension of phonetic interpretation dimension.

One of the most explicit model of the phonetic interpretation of lexical prosodic patterns is that of Aubergé [1], who proposes a lexicon of multiparametric contour or movement prototypes which are concatenated and superposed at different hierarchical levels to generate prosodic patterns paired with a hierarchy of syntactic units. Like Steedman's model, Aubergé's approach is somewhat dependent on the formal peech domain selection required for speech synthesis.

1.3. The Metalocutionary Principle (MP)

MP: *Prosodic patterns are semantically interpreted partly in terms of pragmatic discourse control functions and partly in terms of a metalocutionary semantics of focussing and configuring lexico-grammatical locutions.*

The MP is characterised rather like the SP as follows: $P = construction_{pros}(p_1, ...p_n)$,
$< propPros_{sem}(P),\ propPros_{phon}(P) > = composition(< propPros_{sem}(p_1),\ propPros_{phon}(p_1) >, \ldots\ < propPros_{sem}(p_n),\ propPros_{phon}(p_n) >$
but where sem (the locution) and $phon$ (the prosody) are both contained in the phonetic domain.

The MP is a special case of the SP, but the central one for prosody. The traditional "grammatical" or "diacritic" functions of intonation are the classic examples of the metalocutionary semantics of prosodic patterns. The meaning of an accent is fundamentally deictic: "listen, *this* item is in focus". The meaning of initial and final boundary tones is likewise fundamentally deictic: "listen, from *this* item to *this* item is a sense unit" [4]. The notions *focus* and *sense unit* are left undefined in the present context and require further explication.

It is well-known that scope of this metadeictic semantics for prosody is variable: an accent may "point" to a specific syllable (perhaps even syllable constituent) in contrastive stress, or to a word, a phrase, or even to a longer utterance constituent.

2. Integrated Lexicon modelling conventions

The three principles, CP, SP and MP are taken as axiomatic, but they are empirical, falsifiable generalisations about the Integrated Lexicon model (ILEX). For instance, the CP is in part falsifiable by the existence of prosodic idioms, as in the intonation of greetings or other fixed expressions, which are compositional in form but not in meaning, or by the existence of prosodic items (such as "call contours") which occur in isolated utterances, not in mid-text. The SP is falsifiable by prosodic items which have no semantic interpretation except distinctiveness, such as lexical stress, lexical pitch accent and lexical tone patterns [9], or perhaps (if they exist) neutral or broad scope accentuation patterns.

From these considerations, basic descriptive requirements can be derived for an Integrated Lexicon:

1. Basic units must be enumerated and interpreted terms of their properties in a finite vocabulary.
2. For complex, partially idiosyncratic lexical items (collocations, idioms) a hierarchy of partial compositionality (or inversely: partial idiosyncrasy) is required.
3. A commonly used formalism for representing lexical information is needed so as to relate prosody to other categories and functions of language and speech.

Further, the three basic structural properties of the Integrated Lexicon itself need to be characterised in terms of the lexical items (here, the prosodic categories such as tones and accents):

Microstructure: the arrangement of types of lexical information associated with a lexical item (as a vector, AV structure, etc.) as its defining properties.

Mesostructure: the type (or default) hierarchy of classes of lexical items, based on generalisations about types of lexical information.

Macrostructure: the arrangement of lexical items as a list, a rank hierarchy (of simplex and complex items), or a network.

3. Microstructure: compound nuclei

3.1. A microstructure hierarchy

The microstructure of a lexical item specifies its combinatorial properties (sequential and synchronous valency, and internal structure), its semantic interpretation and its modality interpretation (gestural phonetic, i.e. acoustic or visual). A prosodic lexical item such as the accentual tone pattern 'LH*' has, for example,

1. category (accent), sequential valency properties in terms of cooccurrence with other instances of the same category, boundary tones such as 'LH%', etc.,
2. synchronous valency properties in terms of its cooccurrence with lexically and phrasally stressed syllables such as the metalocutionary alignment relation '*',
3. an internal structure in terms of the component tones 'L', 'H' and the metalocutionary relation '%',
4. a content (semantic-pragmatic) interpretation in terms of theme, backgrounding, mutual belief and contrast,
5. a surface modality interpretation
   (a) in gestural-acoustic terms, correlating with the suprasegmental properties of pitch, intensity and duration pattern, and synchronicity with syllabic structure,
   (b) in gestural-visual terms, correlating with facial gesture (eyebrow movement), head gesture (nodding) and arm and hand movement.

The following outline of a microstruture template for prosodic lexical items is proposed:

$$\begin{bmatrix} \text{TYPE: Prosodic Word} & \\ \text{COMPOSITION:} & \begin{bmatrix} \text{CATEGORY: Lexical\_tone, X}^\frown\text{Y, \_}\circ\text{Y} \\ \text{PARTS: } \{\text{Simplex, Complex}\} \end{bmatrix} \\ \text{INTERPRETATION:} & \begin{bmatrix} \text{SURFACE: Pitch, Duration, Energy; Face} \\ \text{CONTENT: Contrast} \end{bmatrix} \end{bmatrix}$$

The traditional attribute names correspond closely to the attribute names used in Head-driven Phrase Structure Grammar (HPSG) paradigm (most recently [14]):

3.2. A microstructure for rise-fall nuclear accents

An even more HPSG-like microstructure hierarchy can be justfied. For example, a complex prosodic lexical item such as L*H L%, a rise-fall pattern often used to express positive or negative appraisal in English, is assigned a prototypical microstructure as follows (slightly abridged):

```
[ Rise-Fall
  SEMANTICS:  <appraisive>
  PHONETICS:  [broad_bandwidth]
  HEAD:       [ Fall
                SEMANTICS:  <certainty>
                HEAD:       [ Low
                              PHONETICS: [L] ]
                SPECIFIER:  [ High
                              PHONETICS: [H] ] ]
  SPECIFIER:  [ Rise
                SEMANTICS:  <suspense>
                HEAD:       [ High
                              PHONETICS: [H] ] ] ]
```

3.3. A microstructure for a call contour

So-called "call contours" (they have other uses in other languages) have been much discussed during the past half-century, but there is no satisfactory notation for them in current prosodic transcription systems. Phonetically, they have chant-like properties; distributionally, they do not occur in direct sequence with other contours; functionally, they are indicators of phatic, channel-maintaining speech acts [4], [7]. As a starting point, a compositional feature structure with a relational feature value (for a falling relation, sometimes interpreted as downstep) and downward feature inheritance can be represented as a tree (Figure 1).

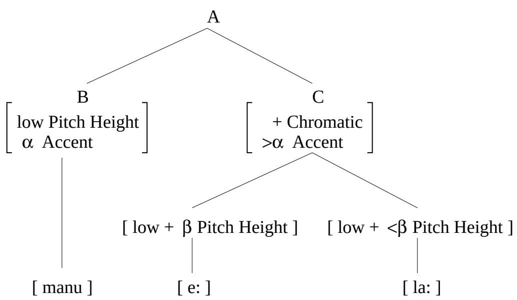

Figure 1: Call contour compositionalilty.

This translates to the following AV structure in more conventional notation, with idiomatic (i.e. holistic, non-compositional) semantics:

```
[ StylisedFall
  SEMANTICS:  <phatic>
  PHONETICS:  [stable, chromatic_interval]
  HEAD:       [ FALL
                HEAD:       [ High
                              PHONETICS: [H] ]
                SPECIFIER:  [ Mid
                              PHONETICS: [M] ] ] ]
```

3.4. Mesostructure outline for nuclear accents

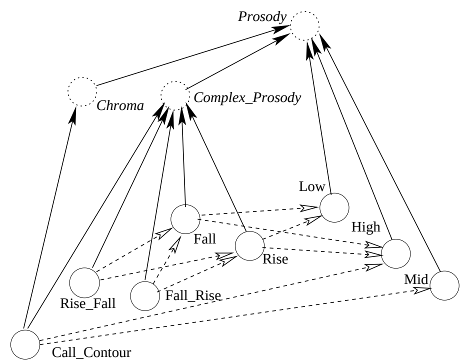

Figure 2: Lexical mesostructure for a set of nuclear patterns.

In the mesostructure, lexical items are grouped into class hierarchies characterised by shared generalisations over microstructures. Such generalisations used to be called *redundancy rules*; in recent lexicalist grammars, *lexical rules* is used, together with a *type inheritance hierarchy* (HPSG), or a *default inheritance hierarchy* (the ILEX model [5]). Figure 2 outlines a mesostructure hierarchy for a set of nuclear contours. Each node is understood as being labelled with microstructure information, idiosyncratic to the item at the lowest level in the hierarchy, and generic at higher levels inherited by the lexical item from its superordinate lexical classes. The compositional macrostructure is indicated by dotted lines.

4. Metalocutionary lexical compounds: prosody-particle pairs

Discourse particles, such as hesitation particles (sometimes called "filled pauses") [3], have characteristic prosodic features [7]. A pitch trace of the German hesitation particle *hm* in context, from the Verbmobil corpus, is shown in Figure 3. The context is

```
cd1_n005k_CR1005:
wie w"are denn der "ahm zehnte Mai ?
```

The first part of the lexical entry defines the compositionality constraints of prosodic *association* (synchronicity, alignment), symbolised by '∘', and *unification*, symbolised by '⊔'); the boxed integer notation of HPSG is used to indicate shared structure constraints:

```
[ discourse–control–sign
  SURFACE     [1] ∘ [2]
  SEMANTICS   [3] ⊔ [4]
  SYNTAX      prosody–particle–pair
  DAUGHTERS   [ HEAD–DTR    [ prosody
                              SURFACE    [1]
                              SEMANTICS  [3] ]
                COMP–DTRS ⟨ [ particle
                              SURFACE    [2]
                              SEMANTICS  [4] ] ⟩ ] ]
```

The microstructure of the compound lexical entry, with a generalised composition operator '⊗' which covers both '∘',

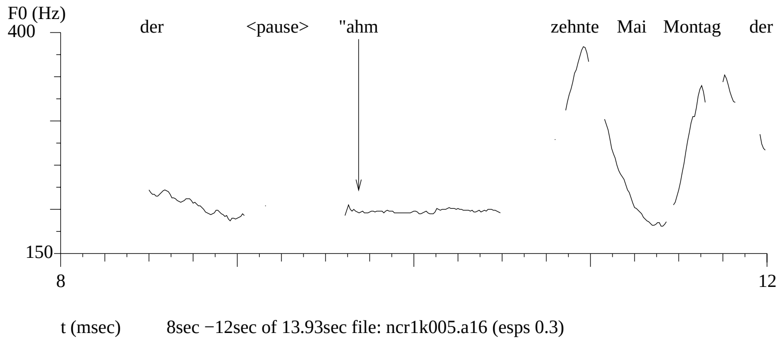

Figure 3: German discourse particle 'ähm' with stylised F_0 trajectory.

and '$\sqcup$', requires both phonetic and semantic-pragmatic attributes:

$$< \text{HES}, \begin{bmatrix} \textit{prosody} & \\ \text{SURFACE} & \left[\text{PHON}\ [\ \text{PITCH}\ \textit{stable}\ ]\right] \\ \text{SEMANTICS} & \left[\text{CC} \begin{bmatrix}\text{MEDIUM} & \textit{vocal}\\ \text{CONTACT} & \begin{bmatrix}\text{FUNCTION} & \textit{phatic}\\ \text{CONTROL} & \textit{feedback}\\ \text{ROLE} & \textit{initiator}\end{bmatrix}\end{bmatrix}\right]\end{bmatrix} \otimes \begin{bmatrix} \textit{particle} & \\ \text{SURFACE} & \begin{bmatrix}\text{PHON} & /\varepsilon\text{:m}/\\ \text{ORTH} & \text{ähm}\end{bmatrix} \\ \text{SEMANTICS} & \begin{bmatrix}\text{QUALIA} & \emptyset\\ \text{INDEX} & \emptyset\\ \text{PLAN} & \textit{hesitation}\end{bmatrix}\end{bmatrix} >$$

5. Conclusion and outlook

An Integrated Lexicon approach to representing the properties of autonomous prosodic lexical items in the lexicon was proposed, with three general principles determining kinds of lexical information. Attribute-value microstructures and a mesostructure were defined for a number of nuclear contours, with a compositional macrostructure relation linking compound lexical items with simplex lexical items.

A number of aspects of this proposal are still fragmentary. Some features of the formalisation are *ad hoc*. An optimal formulation of the semantic microstructure component needs to relate the syntax of locutions and the microstructures of speech act indicators such as speech act verbs, subjective adverbs, and discourse particles, especially modal particles. The locutionary rank hierarchy (syllable, word, sentence, utterance, dialogue), needs to be related in detail to the prosodic hierarchy; earlier feature-based work on prosody (cf. Hirst [8], Selkirk [15] and others) is relevant here. The lexical component needs to receive explicit, computational phonetic interpretation, relating to work of Aubergé or Taylor on parametrised protoypes and to work on Finite State Prosody. The Integrated Lexicon approach provides a fruitful foundation for this work.

6. References

[1] Aubergé, V., 1992. Developing a structured lexicon for synthesis of prosody. In Bailly, G.; Benoît, C., eds. *Talking Machines: Theories, Models and Designs.* Elsevier B.V., 307–321.

[2] Bird, S., 1991. The Intonational Phrase in Sign-based Grammar. Workshop on Declarative Perspectives on the Syntax-Prosody Interface. Utrecht.

[3] Fischer, K., 2000. *From Cognitive Semantics to Lexical Pragmatics: The Functional Polysemy of Discourse Particles.* Mouton de Gruyter: Berlin, New York.

[4] Gibbon, D., 1976. *Perspectives of Intonation Analysis*. Bern: Lang.

[5] Gibbon, D., 1992. ILEX: A linguistic approach to computational lexica. In: U. Klenk, ed., *Computatio linguae.* Beihefte der Zeitschrift für Dialektologie und Linguistik, 32-53.

[6] Gibbon, D., 2001. Finite state prosodic analysis of African corpus resources. *Proc. EUROSPEECH 2001*, Aalborg.

[7] Gibbon, D.; Sassen, C., 1997. Prosody–particle pairs as discourse control signs. In: *EUROSPEECH 97, Proc. 5th European Conference on Speech Communication and Technology*, Vol. I, ed. G. Kokkinakis et al. U Patras: Patras, Greece.

[8] Hirst, D., 1977. *Intonative Features. A Syntactic Approach to English Intonation.* Mouton: The Hague.

[9] Hirst, D.; Di Cristo, A., eds., 1998. *Intonation Systems : A Survey of Twenty Languages.* Cambridge: Cambridge University Press.

[10] Jansche, M., 1998. A two-level take on Mandarin tone. Ivana Kruijff-Korbayová, ed., 1998. Proc. 2rd ESSLI Student Session, 162-174.

[11] Kaplan, R., Kay, M., 1994. "Regular models of phonological rule systems", *Computational Linguistics* 20 (3) 331-378.

[12] Ladd, D. R., 1996. *Intonational Phonology.* Cambridge: Cambridge University Press.

[13] Pierrehumbert, J.; Hirshberg, J., 1989. The Meaning of Intonational Contours in Discourse. In: Cohen, P.R., ed., *Intentions in Communication.* Cambridge (MA) etc.: Bradford, 271-311.

[14] Sag, I. A.; Wasow, T., 1999. *Syntactic Theory: A Formal Introduction.* Stanford: CSLI.

[15] Selkirk, E., 1982. *The Syntax of Words.* Cambridge, MA: MIT Press.

[16] Steedman, M., 1996. Representing discourse information for spoken dialogue generation. *Proc. Int. Symposium on Spoken Dialogue,* ICSLP-96, Philadelphia, Oct. 1996.