1 *Speech Communication in Human Interaction*

1.1 Human Interaction and the *Organon Model*

Humans interact for a variety of reasons:

- for survival and procreation, and for play, which they share with the animal world
- for creating habitats and social bonds, which they basically share with many animals
- for making tools and using them in their daily activities
- for selling and buying, and for business transactions in general
- for establishing, enforcing and observing social and legal codes
- for social contact, phatic communion and entertainment
- for reporting events and issuing warnings
- for instructing and learning
- for asking, and finding answers to, questions of religious belief, of philosophical understanding, of scientific explanation, of historical facts and developments
- for artistic pursuits for eye, ear and mind.

Central to all these human interactions is speech communication, i.e. communication via an articulatory–acoustic–auditory channel (AAA) between a sender and a receiver, supplemented by a gestural–optical–visual channel (GOV). Speech communication is based on cognitive constructs that order the world and human action in space and time. These constructs are manifested in the AAA channel as words with their paradigmatic and phonotactic sound structures, and as syntagmatic organisations of words in utterances. The words and phrase structures are linked to articulatory processes in speech production by a speaker and to auditory patterns in speech perception and understanding by a listener. Speech communication on the basis of such cognitive constructs, of their formal representation, and articulatory and perceptual substantiation,
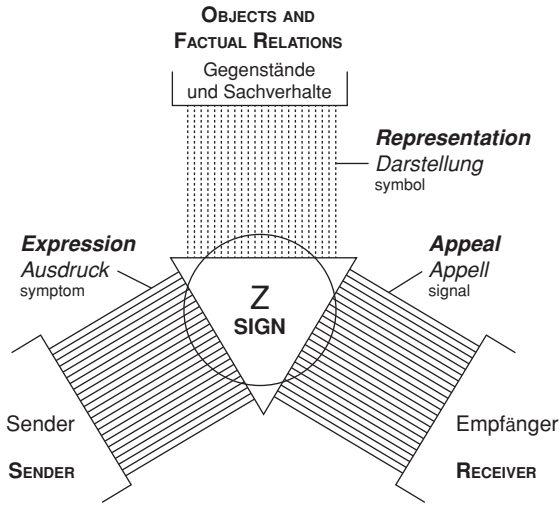
18

Figure 1.1. The *Organon Model* according to Bühler (1934, p. 28), with the original German labels, and their added English translations, of the three relationships, functions and aspects of the linguistic SIGN Z(eichen).

shared by speakers and listeners, performs three basic functions in sender–receiver interaction:

(a) the transmission of *symptoms* relating to the sender's feelings and attitudes in the communicative act

(b) the emission of *signals* by a sender to a receiver to stimulate behaviour

(c) the transmission of *symbols* mapped to objects and factual relations in space and time, constructing the world in communicative acts.

This is Karl Bühler's *Organon Model* (from Classical Greek ὄργανον, 'instrument, tool, organ', after Aristotle's works on logic: Bühler 1934, pp. 24–33; see Figure 1.1), which relates the linguistic sign to the speaker, the listener, and the world of objects and factual relations, in the communicative functions of EXPRESSION (a), APPEAL (b) and REPRESENTATION (c). The objects in the symbolic mapping of (c) are not just concrete things, e.g. 'table', 'mountain', but also include abstract entities, e.g. 'love', 'death', and attributes, e.g. 'redness', 'beauty'. The symbolic mapping to objects and factual relations constitutes *language structure* [*Sprachgebilde*] based on social convention binding individual speakers in their *speech actions* [*Sprechhandlungen*]. This is, in Bühler's terms (1934, pp. 48ff), de Saussure's *langue* versus *parole* (de

Saussure 1922). For the REPRESENTATION function, the human mind devised systems to capture linguistic signs graphically on durable material in order to overcome the time and space binding of fleeting signals through AAA and GOV channels. These writing systems are either logographic, with reference to the symbolic values of linguistic signs, or phonographic, with reference to their sound properties, either syllabic or segmental. The latter, alphabetic writing, was only invented once in the Semitic language family. It conquered the world and became the basis of linguistic study, which has, for many centuries, focused on the REPRESENTATION function in written texts, or on speech reduced to alphabetic writing.

The three aspects of the linguistic sign – sender *symptom*, receiver-directed *signal*, *symbol-to-world* mapping – are semasiological categories, with primary manifestation through the AAA channel, but accompanied in varying degrees by the GOV channel, more particularly for the functions (a) and (b). Bühler made it quite clear that he regarded the three functions of his *Organon Model* as being operative in any speech action at any given moment, but with varying strengths of each, depending on the communicative situation. In rational discourse, REPRESENTATION with *symbol-to-world* mapping dominates; in highly emotional communication, it is the *symptoms* of the EXPRESSION function; in commands on the drill ground, the *signals* of the APPEAL function; a balance of *signals* and *symptoms* occurs in words of endearment or abuse. An aggressive act may be totally devoid of symbolic meaning, as in the reported case of a Bonn student silencing the most powerful market crier in the Bonn fruit and vegetable market, eventually having her in tears, by simply reciting the Greek and Hebrew alphabets loudly with pressed phonation: 'Sie Alpha! Sie Beta! …' (Bühler 1934, p. 32).

The linguistic sign is at the centre of the model and has a direct iconic symptom or signal relationship to the sender or the receiver in EXPRESSION and APPEAL, respectively, and an indirect symbolic relationship to objects and factual relations in REPRESENTATION. The direct or indirect relationships are indicated by plain or dotted connection lines in Figure 1.1. The linguistic sign is encapsulated in a circle encircling all three functions. Superimposed on, and cutting across, this circle is a triangle connecting with the sign's three functions and covering a smaller area of the circle, as well as going beyond the area of the circle with its three edges. The triangle represents Bühler's principle of *abstractive relevance* applied to the phonetic manifestation of the linguistic sign, which is captured by the circle. The triangle contains only the communicatively relevant features of the total of phonetic properties, and at the same time it adds functional aspects in relation to the three communicative functions that are absent from the phonetic substance.

Abstractive relevance is also the basis for Bühler's concept of phonology versus phonetics, which he expounded in his seminal article of 1931 (Bühler 1931), and which Trubetzkoy took over in his *Grundzüge* of 1939. Abstractive relevance means that the total phonetic substance of the instantiation of a linguistic sign is reduced to its functionally relevant phonetic features by an abstractive scaling, not by abstract representation. Thus, phonology comes out of phonetics, phonetics does not go into phonology, contrary to Ladd (2011), who interpreted Prague phonology as an abstraction from phonetics that stopped short of its logical conclusion (cf. Kohler 2013a). The mistake Prague phonology made was not incomplete phonological abstraction from concrete phonetics but the postulation of two disciplines, phonology and phonetics, of which the former was furthermore linked with the humanities, the latter with the natural sciences. The reason this happened lies in the methodology of early experimental phonetics at the turn of the twentieth century, where linguistic concepts disintegrated and objective truth was sought in speech curves of various, mainly articulatory, origins and in the numbers derived from them (Scripture 1935, p. 135). This imbalance was put right again by the Prague linguists, who reintroduced the functional aspect into phonetics, which had always been present in the several thousand years of descriptive studies of speech sounds in languages since the invention of alphabetic writing. Bühler's concept of abstractive relevance shows how this dichotomy can be overcome: there is only one science of the sound of speech in human language – it determines the functionally relevant features in speech communication in the languages of the world from the broad array of sound in individual speech acts.

Bühler developed the concept of abstractive relevance in connection with the symbolic mapping of sound markers of the linguistic sign to Objects and Factual Relations in the REPRESENTATION function, especially the sound markers of names (words) assigned to objects. The entire sound impressions of words are not relevant for the differential name-object mappings; only a small number of systematically ordered distinctive sound features are. This is the principal aspect of Prague segmental word phonology incorporated into Bühler's theory of language. Apart from lexical tone and stress, this framework says nothing about prosodic phonology at the level of mapping formal phrasal structures and factual relations in the world. Bühler left a gap in his theory of language, which needs filling in two respects:

- He considers 'musical modulation' at the utterance level in the Indo-European languages to be irrelevant for REPRESENTATION, and therefore free to be varied diacritically in the other two functions, for example, adding an urgent APPEAL to the German phrase 'es regnet'

[it is raining] in order to remind a forgetful person to take an umbrella (Bühler 1934, p. 46). Global unstructured utterance prosodies are seen as EXPRESSION or APPEAL overlays on structured phonematic lexical sound markers in REPRESENTATION. This is incomplete in two respects: prosody can and does map symbolic relations in REPRESENTATION, and it is highly structured in all three functions. In Bühler's time, prosodic research was still in its infancy, so he was not able to draw on as rich a data analysis as we can today.

- The function-form perspective is to map the functions of the *Organon Model*, as well as subfunctions in each, and their formal systems and structures at all linguistic levels, from phonetics/phonology through the lexicon and morphology to syntax. For example, the investigation into QUESTION versus STATEMENT needs to consider word-order syntax, question particles and prosody. In this way, prosody as the acoustic exponent in symbolic phrase-level mapping is treated on a par with other formal means, lexical and structural, and is thus fully integrated into the theory of language and of language comparison.

Bühler saw the gap in his theory and set a goal for further development:

> Let me stress the point once again: these are only phenomena of dominance, in which one of the three fundamental relationships of the language sounds is in the foreground. The decisive scientific verification of our constitutional formula, the Organon Model of language, has been given if it turns out that each of the three relationships, each of the three semantic functions of language signs discloses and identifies a specific realm of linguistic phenomena and facts. That is indeed the case. 'Expression in language' and 'appeal in language' are partial objects for all of language research, and thus display their own specific structures in comparison with representation in language … This is the thesis of the three functions of language in simplest terms. It will be verified as a whole when all three books that the Organon Model requires have been written. (Bühler 1934, p. 31f [2011, p. 39])

He himself concentrated on the REPRESENTATION function, which he indicated in the subtitle of the book. It resulted from an extensive study of the extant literature of Indo-European historical linguistics, with its focus on such topics as the Indo-European case system, deixis and pronouns, anaphora, word and sentence, compound, ellipsis and metaphor, generally dealing with REPRESENTATION, written texts and historical comparison. He was also thoroughly familiar with the Greek philosophers, with modern logic and with the philosophy of language. He especially discussed Husserl's *Logische Untersuchungen* and *Cartesianische Meditationen* in some detail in connection with the concept of

Sprechakte [*speech acts*], in which a speaker confers specific discourse-driven meanings to words of a language, and which are distinguished from *Sprech-handlungen* [*speech actions*], the unique *hic et nunc* utterances by individuals. He also took de Saussure (1922), and especially Gardiner (1932) and Wegener (1885), into account, who added the linguistic expert perspective to what Bühler contributed as a psychologist working with language. The study of language is a study of creative actions, not of a static linguistic object, because language users interact through speech actions in communicative speech acts by means of a *Sprachgebilde* [*language structure*] to create *Sprachwerke* [*language works*]. This naturally led to the *Organon Model* and to looking beyond Representation.

1.2   Deictic and Symbolic Fields in Speech Communication

In addition to the *Organon Model*, Bühler proposed a two-field theory of speech communication: the pointing or *deictic* field and the naming or *symbolic* field. The deictic field is one-dimensional, with systems of deictic elements that receive their ordering in contexts of situation. In a pointing field, a speaker sets the sender *origo* of *hic-nunc-ego* coordinates, which position the speaker in space and time for the communicative action. Within the set coordinates, the sender transmits gestural and/or acoustic signals to a proximate or distant receiver. These signals point to the sender, or to the receiver, or to the world of objects away, or far away, from (the positions of) the sender and the receiver. Receivers relate the received sender-, receiver- or world-related signals to their own *hic-nunc-ego* coordinates to interpret them. The understanding of their intended meanings relies on material signal properties that guide the receiver through four different pointing dimensions: *here* or *hic* deixis; *where-you-are* or *istic* deixis; *there or illic* deixis; and *yonder* deixis.

On the other hand, the symbolic field in its most developed *synsemantic* form is a field where linguistic signs do not occur primarily in situational but in linguistic contexts. It is two-dimensional, comprising systems of signs for objects and factual relations, a lexicon, and structures, a syntax, into which the systemic units are ordered. Another symbolic field is the one-dimensional *sympractical* field, which contains systems of signs that are situation-related in an action field, rather than being anchored in linguistic context.

1.2.1   Deictic Field Structures
In deictic communication, the sender creates a situational field in space and time by using optical and acoustic signals in relation to the sender's

hic-nunc-ego coordinates, and the receiver decodes these signals with reference to the receiver's position in the created communal space-time situation. If the signals are optical, they are gestures, including index finger or head pointing, and eye contact. If they are acoustic, they include linguistic signs, deictic particles, demonstrative and personal pronouns, which function as attention signals. These signals structure the deitic field with reference to the four pointing dimensions. For each dimension, the relation may be *at*, *to* or *from* the reference, as in the Latin deictic signs 'hic', at the sender, 'huc', to the sender, 'hinc', from the sender; 'istic', at the receiver, 'istuc', to the receiver, 'istinc', from the receiver; 'illic', at a third-person place, 'illuc', to that place, 'illinc', from that place. Yet pointing in a situational field is always meant for a receiver, even if there are no specific receiver-deictic signals. The linguistic signs receive their referential meaning through the situational dimensions. 'here', 'I', 'yours', 'that one over there' are semantically unspecified outside the *hic-nunc-ego* coordinates of the deictic field. Languages differ a great deal in the way they structure the deictic field with deictic linguistic signs. Latin provides a particularly systematic place-structure deixis. Linguistic deixis signs are not only accompanied by gestures, but also by acoustic signals pointing to the sender or the receiver.

1.2.1.1   *here* or *hic* Deixis
Hic deixis signalling points to the sender in two ways, giving (1) the position and (2) the personal identification of the sender.

(1) Position of the Sender   When speaker B answers 'Here' from a removed place after speaker A has called out, 'Anna, where are you?', the deictic 'here' is defined within sender B's coordinates but remains unspecified for receiver A unless the acoustics of the uttered word contain properties pointing to the sender's position in space, or, in the case of potential visual contact between sender and receiver, are accompanied by a gestural signal of a raised hand or index finger. The acoustic properties of signal energy and signal directionality give A a fair idea of the distance and the direction of B's position in relation to A's coordinates, indicating whether the sender is nearby, e.g. in the same room or somewhere close in the open, but outside A's visual field, or whether B is in an adjoining room, or on another floor or outside the house. From their daily experience with speech interaction, both speaker and listener are familiar with the generation and understanding of these sender-related pointing signals.

A different variety of this *hic* deixis occurs in response to hearing one's name in a roll call, which depends on visual contact for verification. Raising arm and index finger and/or calling out 'Here'/'Yes' transmits the sender's position and personal (see (2) below) coordinates.

> *(2) Personal Identification of the Sender* A speaker B, waiting outside the door or gate to be let in, may answer 'It's me' in response to a speaker A asking 'Who is there?' over the intercom. This *hic* deixis is only understandable if A has a mental trace of B's individual voice characteristics. Their presence in a pointing signal allows the correct interpretation of an otherwise unspecified 'me'. Again, speakers and listeners are familiar with these material properties of individualisation in speech interaction.

1.2.1.2 *where-you-are* or *istic* Deixis

Istic DEIXIS signalling points to the receiver. Bühler thought that, contrary to the other types of deixis, there are no specific systematic pointing signals for *istic* DEIXIS, although he lists a few subsidiary devices on an articulated sound basis, such as 'pst', 'hey', 'hello', or the reference to the receiver by 'you', or by the personal name, accompanied by index finger pointing and/or head turning to the person to establish eye contact.

However, an examination of the various occurrences of *istic* DEIXIS, and of the pitch patterns associated with them in English and German, shows up a specific melodic device that is characteristic of utterances pointing to the receiver, and that differs from pitch patterns used in other types of deixis and in the synsemantic symbolic field of speech communication. It is level pitch stepping up or down, or staying level, as against continuous movement (see 2.14). Continuous pitch patterns form a system of distinctive differences for coding REPRESENTATION, APPEAL and EXPRESSION functions in speech communication in a particular language. They fill speech communication with sender-receiver-world content. On the other hand, stepping patterns function as pointing signals to the receiver to control sender-receiver interaction; they do not primarily fill it with expression of the speaker, attitudes towards the listener and representation of the world. The referential content is predictable from the discourse context, and the sender shares an established and mutually acknowledged routine convention with the receiver. These acoustic patterns of *istic* DEIXIS may occur interspersed in speech communication at any moment to initiate, sustain and close speaker–listener interaction, with two different functions, either to control connection with a receiver or to induce specific action in a receiver. In all the varieties of *istic* DEIXIS found in English and

German, which will be discussed as subcategories of an Appeal function in 4.1, stepping pitch patterns operate as such receiver-directed control signals. When they are replaced by continuous contours in the same verbal contexts they lose the simple pointing control characteristic and become commands, expressive pronouncements and informative statements in acts of speech communication.

1.2.1.3   Proximate and Distant Pointing: *there* or *illic* Deixis and *yonder* Deixis

In order to point away from sender and receiver to objects in a proximate or a distant pointing field, arm and finger gestures are used as the standard signal. Demonstrative pronouns and position adverbs, such as 'that' – 'yon', 'there' – 'yonder'/'over there' in English, or 'der' – 'jener', 'da' – 'dort', 'dort' – 'dort drüben'/'jenseits' in German are linguistic signs used for pointing in sympractical usage, accompanied by gesture, but they also operate anaphorically in a synsemantic field. The distinction between proximate and distant positions in a speaker's pointing field coordinates is less clearly defined than the one between the positions of sender and receiver. Languages do not always have a stable, formally marked system of proximate and distant position adverbs and demonstrative pronouns. Even English 'yon' and 'yonder' are literary and archaic outside dialectal, especially Scottish, usage, and German 'da' versus 'dort' are unstable in their position references. Speakers use phrase constructions instead to define different field positions, for example 'over there' in English and 'dort drüben' in German, or they define distant positions in relation to landmarks. As regards signalling proximate and distant positions by gesture, stretching out arm and index finger in the direction of an object may be used for the former, an upward-downward arm–index finger movement for the latter.

Signalling in a pointing field may combine a deixis gesture to objects with a deixis gesture to the receiver. I recently observed an instance of this. I had just got some cash out of an ATM but was still close to the machine when another customer approached to use it. He turned his head towards me and, with his far-away arm and index finger, pointed to the machine, asking 'Fertig?' [Finished?] (with high-level pitch). He identified the object he wanted to use after me and, by looking at me, identified me as the receiver of his object pointing and his enquiry, which he spoke with the acoustic stylisation of *istic* Deixis, a high-level pitch signalling 'May I use this machine?' This is a Deixis Appeal, different from the Question Appeal 'Is it true that you are finished with the machine?' (see 4.1, 4.2). The response may be 'Ja, bitte' or just 'Bitte' [(Yes,) go ahead], which is impossible in the Question context.

Now let's visit a Scottish pub to illustrate the whole gamut of communicative interaction that is possible with ordering beer, from a synsemantic description to mere gesture. One may give the order 'A pint of Caledonian 80/- please', with continuous pitch. In this situation, it is quite clear that it means 'I want to buy a pint of heavy draught beer sold under the Caledonian Company's trademark', but nobody would use this synsemantic description. The order may be shortened to 'A pint of Caley 80, please', again with continuous pitch. Or one may just point to the label on a draught pump and say 'Pint, please', with a high-level pitch pattern, to induce the receiver to act. This is accompanied by turning one's head towards the barman, to establish gestural contact. Or the customer may point towards the pump with one hand and hold up the other hand with fingers raised according to the number of pints wanted. In a pub near the Tynecastle Hearts football stadium in Edinburgh, called 'The Diggers', because it used to be frequented by gravediggers from a nearby cemetery, this gesturing can be further reduced to mere finger raising, because the barmen know that regulars drink one of their fourteen types of heavy, of course by the pint, and they also know who drinks Caley 80.

The speech versions of this pub order are self-sufficient sympractical signs in their own right, not ellipses of a synsemantic structure 'Sell me a pint of Caledonian 80/- heavy draught beer, please.' In such a sympractical communication field the transmission of symbolic meaning through speech is reduced to the minimum considered necessary by the communicators; the action field and the situation supply the referential meaning, and the EXPRESSION and APPEAL functions are of secondary relevance. They may come in when speakers do not get what they want. Interaction still works when gestures take over altogether. In this case, the question of an ellipsis simply does not arise, which also casts doubt on any attempt to derive the linguistically reduced forms from fully elaborated ones.

1.2.2   From Sympractical Deixis in Situations to Synsemantic Symbols in Contexts

A sender communicating with a receiver may establish a *hic-nunc-ego origo* in a deictic field relating to their actual situation. In its simplest form, communication is just by gesture or by gesture accompanying *sympractical* speech, or only by *sympractical* speech pointing to the situation that both sender and receiver are connected with (e.g. 'mind the gap'/'mind your step' announcements on the London Underground/at Schiphol Aiport, see 4.1.3(2)). The pointing in this *sympractical* speech may be done by deictic particles and pronouns with or without finger gesture, as in 'The flowers over there.' Direct pointing by

gesture and/or deictic words may be removed in *synsemantic* place description in relation to the *origo*, as in 'The flowers are on the table at the window in the back room upstairs.' But there is still some pointing in relation to the position of the sender's *origo* in such utterances, because they are only intelligible with reference to the situation both sender and receiver are related to, and they presuppose the receiver's awareness of, and familiarity with, the locality.

In developing talk, a speaker may move the *hic-nunc-ego origo* from the *actual* sender-receiver situation to a place and time in memories and imagination, and relate symbols to this new *origo* position, thus creating a virtual deictic field, which Bühler calls *Deixis am Phantasma* [*Phantasma Deixis*] (Bühler 1934, pp. 121ff). In Indo-European languages the same deictic signs are used as for pointing in actual situations ('this (one)', 'that (one)', 'here', 'there'; German 'dieser', 'jener', 'der(jenige)'), supplemented by position and time adverbs and conjunctions. This pointing in displaced virtual situations is found in narrating fairy tales:

> *Es war einmal* ein kleines süßes Mädchen … *Eines Tages* sprach seine Mutter zu ihm: … *bring das der Großmutter hinaus…*

> *Once upon a time* there was a dear little girl … *One day* her mother said to her: … *take this to your grandmother…*

or in storytelling of past or future events:

> *After a five-hour climb* we arrived *at the top of Ben Nevis*. *Here* we first of all had a rest. *Then* we dug into our food. And *when* the fog lifted, we were rewarded by the most spectacular view of the Highland scenery *around* and *below* us.

or in giving direction:

> You take the road *north out of our village*. *When* you get to a junction *turn right*, *then the first left*. You continue *there for about a mile. Then* the castle will come into view.

Communicative action changes completely when symbols are anchored in the context of linguistic structure and are freed from situations. Let's assume that on 3 April 2005, the day after Pope John Paul II died, a passenger on a New York subway train says to the person beside him, 'The Pope's died', referring to what he has just read in the paper. This statement is removed from the place of the communicative situation between sender and receiver: it might have been made anywhere around the world (in the respective languages), but it is still linked to the time when the speaker makes it. In a proposition like 'Two times two is four' the time link is also severed. This is the self-contained

synsemantic use of symbols in a symbolic field to refer to objects and factual relations, valid at all times and places, in statements of mathematics, logic and science.

As regards the intonation of such sentences in oral communicative actions, the occurrence of stepping pitch is all the more likely the stronger the sympractical deictic element. Synsemantic sentences have continuous pitch, rising-falling centered on 'Pope', and on the second 'two' in the above examples. If '↓The ↑Pope's died' is spoken with upward-stepping pitch, it may, for example, come from a newspaper seller in the street attracting the attention of people passing by ('Buy the paper, and read more about the news of the Pope'), i.e. a receiver-directed signal puts the synsemantic sentence into a pointing situation. Similarly, when saying the times tables by rote, for instance teacher-directed in class, an upward-stepping pattern will be given to the synsemantic sentence: '↓One times two is ↑two. ↓Two times two is ↑four. ↓Three times two is ↑six…', or shortened to '↓One two ↑two. ↓Two twos ↑four. ↓Three twos ↑six …'

There is another, quite different way of introducing pointing into the synsemantic symbolic field: anaphora (Bühler 1934, pp. 121ff, 385ff). It reinforces reference to the symbolic context because pointing occurs with backward or forward reference to the internal structure of developing talk in the symbolic field, not with reference to the external situation: the symbolic (linguistic) context functions as the pointing field. In Indo-European languages, the exponents are again the same deictics as for pointing in situations, supplemented by position and time adverbs, conjunctions, relative and third-person personal pronouns, fully integrated into the case system and syntax of the language. For examples from German illustrating the distinction between external situation and internal anaphoric pointing, see Abraham (2011, pp. xxiiff).

1.3   From Function to Form

1.3.1   *Bühler and Functional Linguistics of the Prague School*

Bühler places language functions at the centre of his theory of language and then looks at their mapping with linguistic form, for example in the discussion of the Indo-European case system as a formal device for representing objects and factual relations of the world with symbols in a symbolic field (Bühler 1934, pp. 249ff).

The model is thus eminently suited as a theoretical basis for a function-form approach. The notion of function has played a role in many structural theories of language that ask about the acts language users perform with the

formal tools. Functional theories of grammar strive to define these functions and subsequently relate them to the structural carriers. The most elementary function is the differentiation of representational meaning, in its simplest form in functional phonetics. The Prague School linguists were the first to develop functional structuralism, starting with phonology, based on the principle of the distinction of lexical meaning, rather than on the principle of complementary distribution, as in American behaviourist structuralism.

Under the influence of Bühler's *Organon Model*, Trubetzkoy (1939, pp. 17ff) complemented the phonology of the REPRESENTATION function ('Darstellungsphonologie') by phonologies of the EXPRESSION and APPEAL (conative) functions ('Ausdrucks- und Kundgabephonologie'), which he did not always find easy to separate, and which, following Prague systematising, he allocated to a new discipline, called 'sound stylistics' ('Lautstilistik'), with two subsections. He subsumed vocalic lengthening, as in 'It's **won**derful!', and initial consonant lengthening, as in 'You're a **b**astard!', under the APPEAL function, because he maintained that the speaker signals to the listener to empathise with his/her feelings. Isačenko (1966) rightly criticised this solution as unacceptable psychologising and allocated such data to the EXPRESSION function, which I do likewise. Mathesius (1966) extended the functional perspective to lexical and syntactic form (beside accentuation and intonation) for INTENSIFICATION and for INFORMATION SELECTION AND WEIGHTING. Contrary to general usage, he called the latter *emphasis*. Since this term is used with a wide array of signification I shall avoid it altogether and refer to the two functions by the above pair of terms. An INTENSIFICATION scale will be incorporated into the *Organon Model* as the EXPRESSIVE LOW-TO-HIGH KEY function (see Chapter 5).

Jakobson (1960) took up the three functions of Bühler's model as *emotive*, *conative* and *referential*, oriented towards *addresser*, *addressee* and *message referent*. He derived a *magic, incantatory* function from the triadic model as a 'conversion of an absent or inanimate "third person" into an addressee of a conative message' (p. 355). Prayer comes under this heading. And he added another three functions (pp. 355ff):

- *phatic* serving to establish, prolong or discontinue communication: 'Can you hear me?' 'Not a bad day, is it?' – 'It isn't, is it, could be a lot worse' (an exchange between two hikers meeting in the Scottish hills on a foggy, drizzly day)
- *poetic* focusing on the message for its own sake: rhythmic effects make 'Joan and Margery' sound smoother than 'Margery and Joan';

the poetic device of paronomasia selects 'horrible' instead of 'terrible' in 'I **h**ate **h**orrible **H**arry'

- *metalingual*, language turning back on itself: 'What is a sophomore?' – 'A sophomore means a second-year student.'

Jakobson gives the following linguistic criteria for the poetic and metalingual functions:

> We must recall the two basic modes of arrangement used in verbal behavior, *selection* and *combination* … *The poetic function projects the principle of equivalence from the axis of selection into the axis of combination.* Equivalence is promoted to the constitutive device of the sequence. In poetry one syllable is equalised with any other syllable of the same sequence; word stress is assumed to equal word stress, as unstress equals unstress … Syllables are converted into units of measure, and so are morae or stresses … in metalanguage the sequence is used to build an equation, whereas in poetry the equation is used to build a sequence.

Jakobson's additional communicative functions are an extension to Bühler's theory of language, but they are not on a par with the three functions of the *Organon Model*; rather, they are functions within the domains of the sender, the receiver and the referent. The phatic function is clearly receiver-directed and constitutes one type of signalling. The metalingual function belongs to the domain of objects and factual relations, and constitutes the essence of a symbolic speech act. The poetic function is not a function in the sense of the other two, i.e. of communicative action between a sender and a receiver. It is a device characterising a speech act or a language work. As such, it may have an aesthetic function to give sensuous pleasure, or a GUIDE function to increase intelligibility, or a rhetorical function to persuade, as in advertising, in all cases in the domain of the receiver. In the example of paronomasia given above, the poetic device has a speaker-focused EXPRESSION function, which it may also have in reciting lyrical poetry.

Garvin (1994, p. 64), in discussing Charles Morris's three branches of semiotics – syntactics, semantics, pragmatics – notes:

> In Bühler's field theory … variants [of structural linguistic units] can be interpreted in terms of the field-derived properties of the units in question. In the Morrisian schema, I do not seem to be able to find a real place for this issue … None of this, of course, means that I object to 'pragmatics' as a label of convenience for the discussion of certain of the phenomena that, as I have repeatedly asserted, Bühler's field theory handles more adequately, I only object to giving theoretical significance as a separate 'level' or 'component' … the foundation of Bühler's theory is the … Gestalt-psychological notion

> of the figure–ground relation. Morris's foundations, on the other hand, are admittedly behaviorist … There is no doubt about my preference for the Gestaltist position … It is interesting to note that many of the linguists who have arrived at a total rejection of the behaviorist bases of descriptivist linguistics nevertheless have come to use the Morrisian schema, at least to the extent of accepting a pragmatics component for explaining certain phenomena.

In full agreement with Garvin's dictum, I also follow Bühler's theory of language. Building the theory, and the empirical analysis, of language on the *Organon Model* can immediately dispense with all the subdivisions of the field of speech science into separate disciplines, phonology versus phonetics, phonology versus sound stylistics, linguistics versus paralinguistics, pragmatics versus syntax and semantics, and relate units and structures across all linguistic levels of analysis to axiomatically postulated functions in speech communication – functions in the domains of Sender, Receiver and Referent, such as QUESTION, COMMAND, REQUEST, INFORMATION SELECTION AND WEIGHTING, INTENSIFICATION. In moving from these functions to the linguistic signs in their deictic and symbolic fields, speech science can capture all the formal phonetic, phonological and linguistic aspects related to them.

1.3.2   Halliday's Functional Systemic Linguistics

A few words need to be said about another, more recent functional framework that is also rooted in the European linguistic tradition, more particularly J. R. Firth's enquiry into systems of meaning (Firth 1957): Michael Halliday's *Systemic Functional Linguistics* (SFL) (Hasan 2009). It is conceived as systemic with reference to paradigmatic choices in language, and also as functional with regard to specific functions that these formal systems are to serve in communication. These functions are called metafunctions, comprising the *ideational* function (*experiential* and *logical*), the *interpersonal* function and the *textual* function. There are correspondences between Halliday's and Bühler's functions but also fundamental differences. Standing in the European tradition, Halliday and Hasan know Bühler's *Theory of Language*, but they do not always represent it correctly. Hasan (2009, p. 19) says:

> Bühler thought of functions as operating one at a time; further, his functions were hierarchically ordered, with the referential as the most important. The metafunctions in SFL are not hierarchised; they have equal status, and each is manifested in every act of language use: in fact, an important task for grammatics is to describe how the three metafunctions are woven together into the same linguistic unit.

> The concept of 'function', when used in SFL with reference to the system of language as a whole, is critically different from the concept of 'function', as applied to a speech act such as promising, ordering, etc., or as applied to isolated utterances à la Bühler (1934) for the classification of children's utterances as referential, conative or expressive. SFL uses the term 'metafunction', to distinguish functions of langue system from the 'function' of an utterance.

As regards the first quotation, the discussion in this chapter will have made it clear that Bühler's three functions in the *Organon Model* do not operate one at a time, and they are not hierarchically ordered. His linguistic sign has the three functions of EXPRESSION, APPEAL and REPRESENTATION at any given moment, but depending on the type of communicative action their relative weighting changes. In the *Theory of Language*, he puts particular emphasis on the representational function, because this is the area linguistics had been dealing with predominantly during the nineteenth century and up to his time, and he felt a few principles that were generally applied needed to be put right.

The second quotation shows the reason for the misunderstanding. The fundamental difference between the two models is not that Halliday takes a global view of the system of language and Bühler refers to speech actions in isolated utterances. The difference is between a descriptive product model of language in SFL (Bühler's *Sprachwerk*), and a communicative process model of speech actions, which looks at communicative functions between speakers and listeners in speech interaction (Bühler's *Sprechhandlungen*). It is the difference between the linguist's versus the psychologist's view of speech and language. Halliday asks 'How does language work?', whereas Bühler asks 'How do speakers and listeners communicate about the world with linguistic signs in deictic and symbolic fields?' The *Organon Model* is system-oriented, not restricted to utterances, although the functions surface in utterance signals. Halliday's *interpersonal* function is part of all three *Organon* functions: social aspects of the speaker's expression, of attitudes and appeals to the listener, and of representation of the factual world. For Bühler, social relationships determine the communicative interaction between speakers and listeners about referents, i.e. they shape the three functions of the linguistic sign. For Halliday and Hasan, the interpersonal metalevel is a function at a linguistic level, the level of sociolinguistics. The two models are thus complementary perspectives; for a phonetician the process model is particularly attractive because it allows the modelling of speech communication in human interaction.

1.3.3   *Discourse Representation Theory*

More recent language theories have sprung up from logical semantics incorporating context dependence into the study of meaning. A prominent representative of this dynamic semantics is *Discourse Representation Theory* (DRT), developed by Kamp and co-workers (Kamp and Reyle 1993) over the past two decades. Utterances are regarded as interpretable only when the interpreter takes account of the contexts in which they are made, and the interaction between context and utterance is considered reciprocal. 'Each utterance contributes (via the interpretation which it is given) to the context in which it is made. It modifies the context into a new context, in which this contribution is reflected; and it is this new context which then informs the interpretation of whatever utterance comes next' (p. 4). This has resulted in moving away from the classical conception of formal semantics and replacing its central concept of truth by one of information: 'the meaning of a sentence is not its truth conditions but its "information change potential" – its capacity for modifying given contexts or information states into new ones' (Kamp, Genabith and Reyle 2011, p. 4). Anaphoric pronouns referring back to something that was introduced previously in the discourse are the most familiar and certainly the most thoroughly investigated kind of context dependence within this framework.

At first sight, this paradigm looks very similar to Bühler's, and, as its proponents and followers would maintain, is far superior because it is formalised, thus testable, and eminently suited to be applied to the automatic analysis of appropriately tagged corpora. But closer inspection reveals that the two are not compatible. DRT talks about utterances in context but means sentences in textual linguistic environments. However, it is *speech actions* that occur in everyday communication, and they occur not only in synsemantic contexts but, first and foremost, in contexts of situation in sympractical fields. Moreover, not all actions subserve information transmission, because there is *phatic communion* (Jakobson 1960; Malinowski 1923), and appeal to the receiver as well as expression of the sender, where referential meaning is subordinate to social and emotive interaction. DRT would need new categories and a change of perspective, going beyond information structure in texts, to provide explanations for exchanges by speech and gesture, such as the ones experienced on a Kiel bus or in a Scottish pub (cf. Introduction and 1.2.1.3). Here is another set of possible speech actions that illustrate the great communicative variety beyond information exchange in synsemantic text fields:

> I am about to leave the house to go to work, putting on my coat in the hall. My wife is in the adjoining open-plan sitting-room. She briefly looks out of the window and calls to me 'It's raining', with a downstepping level pitch pattern

on 'raining' (see 4.1), to draw my attention to the need to take protection against the weather. I thank her for warning me, grab my umbrella, say 'See you tonight' and leave.

After I have gone, she calls her sister in Edinburgh, and, following their exchanges of greetings, she goes on to talk about the weather, inevitable in British conversation, and asks, 'What's your weather like?', not because she wants to get meteorological information but as an interactional opening. She gets the answer 'It's raining', with a continuous, low falling pitch pattern across the utterance, suggesting 'What else do you expect?' This is followed by a reference to the Kiel weather and then by an appraisal that the recent terrible flooding in the North of England was much worse, so there is really no reason to complain. After this ritual, the two sisters exchange information about family and friends for another half hour, the goal of the telephone call.

After coming off the phone, she switches the radio on to get the 11 a.m. regional news. At the end, the weather forecast reports 'In Kiel regnet es heute' [It is raining in Kiel today]. This is now factual weather information, located in place and time, intended for an anonymous public, therefore removed from interaction between communicators, and since the individual recipient had looked out of the window, the speech action has no informative impact on her.

Each of these communicative interchanges serves a different, but very useful, communicative goal, with different values attributed to the information conveyed. DRT cannot model this diversity because the differently valued types of information are not simply the result of an incremental development of meaning evolving in linguistic contexts but depend on talk in interaction between communicators in contexts of situation. This fact is addressed by Ginzburg (2012) in the *Interactive Stance Model* (ISM).

1.3.4 Ginzburg's Interactive Stance Model

This is a theory of meaning in interaction that, on the one hand, is based on the DRT notion of dynamic semantics and, on the other, incorporates two concepts from Conversation Analysis (Schegloff, Jefferson and Sacks 1977) and from psycholinguistics (Clark 1996): *repair* and *grounding* of content in the communicators' common ground through interaction in contexts. Ginzburg defines the goal of his semantic theory as 'to characterize for any utterance type the contextual update that emerges in the aftermath of successful exchange and the range of possible clarification requests otherwise. This is, arguably, the early twenty-first-century analogue of truth conditions' (Ginzburg 2012, p. 8). This means that an adequate semantic theory must model imperfect communication just as much as successful communication. Besides giving meaning to

indexicals 'I', 'you', 'here', 'there', 'now' through linguistic context in dynamic semantics, non-sentential units, such as 'yes', 'what?', 'where?', 'why?' etc., and repeated fragments of preceding utterances must receive their meanings through the interactive stance in contexts of situation. These are ideas that have been proposed, in a non-formalised way, by Bühler (1934) and Gardiner (1932), as well as by Firth (1957) and his followers in Britain to this day, e.g. John Local and Richard Ogden. None of this literature is cited, no doubt because it is considered outdated and surpassed by more testable and scientific models. However, careful study of the ideas of both camps reaches the opposite conclusion.

Ginzburg's theoretical proposition is that 'grammar and interaction are intrinsically bound' and that 'the right way to construe *grammar* is as a system that characterizes types of talk in interaction' (Ginzburg (2012), p. 349). The pivotal category in this interaction is *gameboards*, one for each participant, which make communicators keep track of unresolved issues in *questions under discussion* and allow for imperfect communication through mismatches. The corollary of the notion of the personal gameboard is that participants may not have equal access to the common ground, and contextual options available to one may be distinct from those available to the other(s). Ginzburg illustrates this with a constructed example of dialogue interaction under what he terms the Turn-Taking Puzzle (p. 23).

a.   A: Which members of this audience own a parakeet? Why?
                  (= Why own a parakeet?)
b.   A: Which members of this audience own a parakeet?
         B: Why? (= Why are you asking which members of this audience
                  own a parakeet?)
c.   A: Which members of this audience own a parakeet? Why am I
                  asking this question?

He explains the different meanings accorded to 'why' in the three contexts by referring them to who keeps, or takes over, the turn. 'The resolution that can be associated with "Why?" if A keeps the turn is unavailable to B were s/he to have taken over, and vice versa. c. shows that these facts cannot be reduced to coherence or plausibility – the resolution unavailable to A in a. yields a coherent follow-up to A's initial query if it is expressed by means of a non-elliptic form.'

These constructed dialogues are problematic, because they lack a sufficiently specified context of situation and violate rules of behavioural interaction beyond speech, and their interpretation by reference to turn-taking is flawed. The reference to 'members of *this audience*' in a book on the Interactive Stance indicates that the speaker must be contextualised as addressing, and interacting

with, a group attending a talk, not as establishing contact for interaction with one or several individuals. Thus B, who is an individual that has not been addressed individually, will not call out from among the audience with non-sentential 'Why?' to ask why the speaker addressed the group with that question. There are three possible reactions from the audience. (1) There is a show of hands by those members who have a parakeet. (2) There is no gestural or vocal response, because nobody in the audience has a parakeet. (3) There may be a call from an obstreperous young attendee, something like 'What the heck are you asking that for? Get on with your subject.' Just as A did not establish interaction with individual members of the audience, speaker B in (3), in turn, does not intend to interact with A, but opposes interaction by refusing to answer A's question.

In response to the reactions, or the lack of a reaction, from the audience, A may continue with 'Why am I asking this question?' (in (2) after pausing for a couple of seconds). In all these cases, A starts a new turn, after a gestural turn from the audience in (1), after a speech turn from an individual member in (3) and after registering absence of a response in (2). A produces an *interrogative* form that is no longer a QUESTION because it lacks the APPEAL to somebody else to answer A's question. It actualises the content of a potential question that the members of the audience may have asked in (1), and particularly in (2), and did ask in (3). This is a QUESTION QUOTE (see 4.2.2.7). Since it is not a QUESTION APPEAL it cannot be reduced to the bare *lexical interrogative*, which presupposes the APPEAL function, and it has falling intonation. In German, the QUESTION QUOTE would be realised by *dependent-clause syntax* '(Sie mögen sich fragen) Warum ich diese Frage stelle?', instead of the *interrogative syntax* 'Warum stelle ich diese Frage?' The latter (as well as its English syntactic equivalent) has two communicative meanings: (a) A appeals to receivers to give an answer why they think A asks the question; (b) it is the speaker's exclamatory expression 'Why on earth am I asking this? (It does not get me anywhere!)' With meaning (b), the interrogative form does not code a question either, since A does not appeal reflexively to A to give an answer to a proposition A is querying. In traditional terminology it would be called a rhetorical question, but in terms of communicative function it is a speaker-centred EXPRESSION rather than a listener-directed APPEAL. Neither (a) nor (b) seem to have a behavioural likelihood in the interaction with an audience. Ginzburg's sequencing of INFORMATION QUESTION and QUESTION QUOTE in one turn in c. may therefore be considered an ill-formed representation of behavioural interaction. Before giving a QUESTION QUOTE to the audience, A must have assessed their reaction to the INFORMATION QUESTION A put to them.

There is a third possibility (c) of contextualising the German and English interrogative forms 'Warum stelle ich diese Frage?' and 'Why am I asking this question?' Here is a possible lecture context (let's assume A is male):

d.   A: I would like to raise a question at the outset of my talk: 'How many of this audience keep parakeets at home?' Why am I asking this question? Well, let me explain. I would like to share experiences of parakeets' talking behaviour with you in the discussion after my talk. So, could I have a show of hands, please. 'Which of you have a parakeet?'

This constructed opening of a lecture illustrates the lecturer's ambivalent function of *reporter to* an audience and *communicator with* an audience. His main function is to report subject matter. In his role as a reporter, the lecturer may raise questions in connection with the topic of the talk, appealing to virtual recipients to give answers. In this reporting role, the lecturer does not enter into interaction with communicators in a real context of situation. He creates a virtual question–answer field in which he enacts interaction between virtual senders and receivers whom he brings to life through his mouth. He treats the audience as external observers of the reporter's question–answer field. This is *question–answer phantasma*, in an extension of Bühler's notion of *Deixis am Phantasma* (see the Introduction and Bühler 1934, pp. 121ff). The lecturer's second function is to enter into an interaction with the audience.

In d., lecturer A is first a reporter, then a communicator. A reports two questions for which virtual receivers are to provide answers in the lecture. The second question is immediately answered by the reporter. These questions differ from the question-in-interaction at the end by being non-interactive. The second question can be a virtual INFORMATION QUESTION with falling intonation, where the reporter enacts the sender and, at the same time, the receiver to give the answer. It may also be a virtual CONFIRMATION QUESTION, with high-rising intonation starting on 'why' (see 4.2.2.4), where the reporter enacts a virtual sender who reflects on his reasons for having asked, and a virtual receiver who is to confirm the reasons in the answer: 'Why am I asking this question really?' Ginzburg's interactive stance excludes both these questions from his context c., but he obviously explains c. in the non-interactive way of d. This problem must have been realised by the reviewer of Ginzburg (2012), Eleni Gregoromichelaki (2013), because she replaced 'this audience' by 'our team' in her discussion of Ginzburg's 'parakeet' example, which is now a question to individual communicators.

Ginzburg's sequencing of a general 'who?' and a more specific follow-up 'why?' INFORMATION QUESTION in one turn of a. is also a behaviourally

ill-formed representation. There must be some response to the first Informa-tion Question before the second one is asked in a new turn. Moreover, if the first question is put to an audience, A needs to select an individual B, or several individuals in succession, for an answer to the second question, because it can no longer be gestural but must be vocal. There is the possibility of a double question, 'Do you own a parakeet and why?', in the opening turn of an inter-action with an individual.

Taking all these points together, there is no compelling reason to associate the attribution of different meanings to non-sentential 'why?' with turn-holding or turn-taking. Ginzburg's explication of this change of meaning in an interaction, with reference to different options available to communicators in their respective turns, is not convincing. He does not provide a sufficiently specified interaction-al setting, does not distinguish between interactions with a group and with an individual, and fails to differentiate Question function and *interrogative* form. Furthermore, he does not acknowledge the occurrence of gestural beside vocal turns, nor of two successive turns by the same speaker, only separated by a pause for the assessment of the interactive point that has been reached. And, *last but not least*, he discusses questions as if they are removed from interaction in spite of their contextualisations. His concept of *interaction* does not model speech *action* in communicative contexts in human behaviour but is derived *post festum* from formal relations in written text, or spoken discourse that has been reduced to writing, or in constructed dialogues dissociated from interaction.

Now let us give Ginzburg's interaction scenario a more precise definition and develop the meanings of the two non-sentential 'why?'s in it.

> [General context of situation
> A famous member of the International Phonetic Association (P) is giving an invited talk to the Royal Zoological Society of Scotland on the subject 'Talking parakeets'. After the introduction by the host and giving thanks for the invitation, P opens his talk.]
>
> a.   P(1):   I suppose quite a few, if not all, of you have a parakeet at home.
>   P(2):   [points to an elderly lady in the front row] What about you, madam? Do you keep one?
>   S(1):   I do. [may be accompanied, or replaced, by nodding] Why?
>   P(3):   Why am I asking you this question. Well, let me explain. I am interested in how owners of parakeets communicate with their pets.
>
> b.   [same precursor as in a., then:]
>   S(1):   I do.
>   P(3):   Why?
>   S(2):   Why? Well, because it keeps me company.

In a., P(2) asks a POLARITY QUESTION (see 4.2.2.2) whether the elderly lady keeps a parakeet in her home, most probably with a falling intonation because the speaker prejudges the answer 'yes'. S(1) answers in the affirmative and asks an INFORMATION QUESTION (see 4.2.2.3), appealing to P to tell her why he asked her. To establish rapport with P, S will use low-rising intonation in both her STATEMENT and her QUESTION. P(3) quotes the content of S's question (see 4.2.2.6), putting it in interrogative form to himself, as a theme for his rheme explanation of his original POLARITY QUESTION. Since the utterance is a factual report, lacking an appeal, the intonation falls. (In German it would be 'Warum ich Ihnen diese Frage stelle', again with falling intonation.) In b., P's POLARITY QUESTION is answered in the affirmative by S, as in a. This is followed by P asking a follow-up INFORMATION QUESTION about the lady's reasons for keeping a parakeet. The intonation may fall or rise, depending on whether P simply asks a factual question or, additionally, establishes rapport with S. This is, in turn, followed by S's CONFIRMATION QUESTION 'Are you asking me why?', with high-rising intonation on the *lexical interrogative* (see 4.2.2.4), in turn followed by her answer.

These examples illustrate communicative steps in a question–answer interaction field, made up of *declarative* and *interrogative* syntactic structures with varying intonation patterns as carriers of STATEMENTS and different types of QUESTION. Different functionally defined question types are bound to the semantic points reached at each step in the interaction and are not exchangeable without changing the semantic context. The crucial issue is that an *interrogative* form does not receive different meanings in different contexts of situation in interaction, as Ginzburg maintains. Rather, the transmission of meaning at different points in interaction necessitates functionally different QUESTIONS, which may be manifested by identical *interrogative* structure. This is the function-form approach proposed in this monograph, which also incorporates an important prosodic component to differentiate between lexically and syntactically identical utterances. Ginzburg's semantic modelling takes an infelicitous turn in three steps:

- He does not recognise *question function* beside *interrogative form*.
- He is forced to locate semantic differentiators in interaction contexts when syntactically identical interrogatives (disregarding utterance prosody), such as 'why', occur with different meanings, and he then incorporates the contexts into the grammar.
- He finally refers the semantic differences of these utterances to their turn-holding or turn-taking positions in dialogue.

The reason Ginzburg tries to resolve the semantic indeterminacy of formal grammar by incorporating context of interaction in it lies in the development of semantics in linguistic theory. The formal component of American Structuralism, as systematised by Zellig Harris (1951, 1960), became the morphosyntactic core of his pupil Noam Chomsky's Generative Grammar (1957, 1965), with a semantic and a phonological interpretive level attached at either end of the generative rule system. Within this generative framework, semantics gradually assumed an independent status, which culminated in DRT. With growing interest in spontaneous speech, the meaning of interaction elements that go beyond *linguistic* context variables had to be taken into account. This led to the inclusion of *situational* context in formal grammar, which became Ginzburg's research goal. It continues the preoccupation with form since the days of structuralism, now with an ever-increasing concern for meaning.

1.3.5   Developing a Model of Speech Communication

To really become an advanced semantic theory of the twenty-first century, the relationship between grammar and interaction would need to be reversed, with a form-in-function approach replacing interaction-in-grammar by grammar-in-interaction. Empirical research within a theory of speech communication can offer greater insight into the use of speech and language than systematising linguistic forms in discourse contexts with grammar-based formalisms. It is a task for the social sciences, including linguistics, to develop a comprehensive Theory of Human Interaction, which contains a sub-theory of Speech Communication, which in turn contains a Grammar of Human Language and Grammars of Languages. Herbert Clark has taken a big step towards this goal by advocating that:

> We must take … an *action approach* to language use, which has distinct advantages over the more traditional *product approach* … Language use arises in joint activities … you take the joint activity to be primary, and the language … used along the way to be secondary, a means to an end. To account for the language used, we need to understand the joint activities [for which a framework of interactional categories is proposed]. (Clark 1996, p. 29)

Influenced by the Language Philosophers Grice (1957), Austin (1962) and Searle (1969), he expanded their theory of meaning in action, *speech acts*, to a theory of meaning in joint activities and joint actions, which accords the listener an equally important role, beside the speaker, in establishing communicative meaning: 'There can be no communication without listeners taking actions too – without them understanding what speakers mean' (Clark 1996, p. 138). However, Clark is first and foremost concerned with *language_u*, the

'language' of language use, which he contrasts with *language*$_s$, the traditional conceptualisation of 'language' as language structure (p. 392). What we need is the incorporation of *language*$_s$ into the theory of speech communication, including the AAA and GOV channels, and a powerful model of fine-graded prosodic systems and structures to signal communicative functions in *language*$_u$.

Since speech and language are anchored in the wider field of human interaction, a communicative approach is the basis of a successful interdisciplinary linguistic science. The seminal thoughts that the psychologist Karl Bühler published on this topic eighty years ago are in no way outdated and inferior to more recent attempts at formalising interaction contexts in grammar. On the contrary, the product approaches of SFL, DRT and ISM, in the tradition of structural linguistics, deal with the formal results of interaction and lose sight of the functions controlling interaction processes, a distinction Bühler captured with *Sprachwerk* [*language work*] versus *Sprechhandlung* [*speech action*]. Since Bühler's theory is little known in the linguistic world, especially among an Anglophone readership, this chapter has given an overview of its main components, to bring them back into the arena of theoretical discussion in formal linguistics and measurement-driven phonetics. I shall pick up Bühler's threads in the following chapters to weave a tapestry of speech communication, and elaborate Bühler's model to a *function* network in human speech interaction to which communicative *form* across AAA and GVO channels will be related. More particularly, I shall provide subcategorisations of the functions of REPRESENTATION, APPEAL and EXPRESSION in Chapters 3, 4 and 5, and integrate prosody, the prime formal exponent of APPEAL and EXPRESSION, into the functional framework of the *Organon Model*. In adding the prosodic level to the analysis of speech interaction, which is largely missing from the formalised context-in-grammar accounts of DRT and ISM, I shall be relying on insights from extensive research on communicative phonetics carried out at Kiel University over the past thirty years.

The communicative model starts from speech functions and integrates with them the production and perception of paradigmatic systems and syntagmatic structures in morpho-syntax, sounds and prosodies. Thus, the functional categories of STATEMENT or QUESTION or COMMAND/REQUEST are separated conceptually and notationally from the syntactic structures of *declarative* or *interrogative* or *imperative*, with distinctive prosodic patterns coding further functional subcategorisations. In German and English, various syntactic structures can be used, with different connotations, of course, to code a COMMAND or a REQUEST:

imperative
with falling intonation for a COMMAND or rising intonation for a REQUEST
Mach (bitte) das Fenster zu!   Shut the window (please)!

interrogative
with falling intonation and reinforced accents for a COMMAND
Machst du endlich das Fenster zu!   Are you going to shut the window!
with rising intonation and default accents for a REQUEST
Würdest du bitte das Fenster   Would you like to shut the window!
zumachen!

declarative
with falling intonation and reinforced accentuation for a COMMAND
Du hast die Tür offen gelassen!   You have left the door open!
Du hast vergessen, die Tür   You forgot to shut the door!
zuzumachen!
Du machst jetzt das Fenster zu!   You are going to shut the window
                                     at once!

Or a QUESTION

interrogative
for a POLARITY QUESTION
Ist er nach Rom gefahren?   Has he gone to Rome?

declarative
with rising intonation or in high register for a CONFIRMATION QUESTION
Er ist nach Rom gefahren?   He's gone to Rome?

Furthermore, within STATEMENT or QUESTION or COMMAND/REQUEST, functional relations between semantic constituents are manifested by syntactic structures between formal elements. Both are enclosed in < >, the former in small capitals, the latter in italics (for some of the notional terminology, see Lyons 1968, pp. 340ff):

<ACTION/OCCURRENCE> <AGENT> <GOAL> <RECIPIENT> <TIME> <PLACE> <MANNER>
<verb> <subject> <direct object> <indirect object> <adverbial or prepositional phrases>

In the *active* versus *passive* constructions of Indo-European languages, <AGENT> is coded by *<subject>* and *<prepositional phrase>*, <GOAL> by *<object>* and *<subject>*, respectively.

<AGENT *subject* > <ACTION *verb*>   <GOAL *dir object*>
<Die Nachbarn>   <verprügelten>   <den Einbrecher>.
<The neighbours> <beat up>   <the burglar>.

<GOAL *subject*>   <ACTION *verb infl*> <AGENT *prepos phrase*> <ACTION *verb uninfl*>
<Der Einbrecher>   <wurde>   <von den Nachbarn>   <verprügelt>.
<The burglar>   <was beaten up>   <by the neighbours>.

The <ACTION>, coded by the unitary *<verb>* 'verprügelten' or 'beat up', may be divided into the semantic dyad <ACTION> <GOAL> coded by the *<verb>* *<direct object>* phrase 'verpassten eine gehörige Tracht Prügel' or 'gave a good beating', making 'Einbrecher' or 'burglar' the <RECIPIENT *indirect object* > of <ACTION> <GOAL>. *Active* can again be turned into *passive*.

<AGENT *subjext*>	<ACTION *verb*>	<RECIPIENT *indir object*>	<GOAL *dir object*>
<Die Nachbarn>	<verpassten>	<dem Einbrecher>	<eine Tracht Prügel>.
<The neighbours>	<gave>	<the burglar>	<a good beating>.

<RECIPIENT *indir object*>	<ACTION *verb infl*>	<AGENT *prepos phrase*>
<Dem Einbrecher>	<wurde>	<von den Nachbarn>

<GOAL *subject*>	<ACTION *verb uninfl*>	
<eine Tracht Prügel>	<verpasst>.	

<RECIPIENT *subject*>	<ACTION *verb*>	<GOAL *object*>	<AGENT *prepos phrase*>
<The burglar>	<was given>	<a good beating>	<by the neighbours>.

Finally, the passive patient construction may be lexical:

<RECIPIENT *subject* >	<ACTION *verb*>	<AGENT *prepos phrase*>	<GOAL *object*>
<Der Einbrecher>	<erhielt>	<von den Nachbarn>	<eine Tracht Prügel>

<RECIPIENT *subject*>	<ACTION *verb*>	<GOAL *object*>	<AGENT *prepos phrase*>
<The burglar>	<got>	<a good beating>	<from the neighbours>.

Another type of proposition centres on an <EVENT>, for instance meteorological events:

<EVENT>	<Time>/<Place>	<Place>/<TIME>
<Es regnet/schneit>	<heute>	<in Paris>.
<It's raining/snowing>	<in Paris>	<today>.

In these cases, both the event and its occurrence are coded syntactically by the impersonal verb construction. But, more generally, the two semantic components are separated in syntactic structure, for instance as *<subject>* and *<verb>*, and German and English may go different ways, for example in:

Zur Zeit ist über Paris ein Unwetter.
<EVENT OCCURRENCE>
<Es stürmt, hagelt, blitzt und donnert>

There is a heavy thunderstorm over Paris right now.

<OCCURRENCE >	<EVENT>
<There are>	<gale-force winds, hail, thunder and lightning>

Before I move on, let me add a word of clarification concerning the difference, and the relationship, between communicative theory and linguistic discovery procedures. It is a well-established, very useful goal in linguistics to

work out the systems and structures of distinctive phonetic sound units that are used to distinguish words in a language, including lexical tone, lexical stress and phonation type in tone, lexical stress and lexical voice quality languages. It is mandatory to base this investigation on the word removed from communicative context in interaction. There is an equally established and useful procedure to work out the morpho-syntactic elements and structures, as well as the accent and intonation patterns that carry distinctive sentential meaning. This puts the sentence removed from communicative context in focus. In the initial analysis stages of an unknown, hitherto uninvestigated language, these phonological and syntactic discovery procedures produce context-free word and sentence representations, which will have to be adjusted as the investigation continues and more and more context is introduced in a series of procedural steps. But it will not be possible to base the phonetic or syntactic analysis on talk in interaction for a long time yet. The procedural product approach makes it possible to reduce a language to writing, and to compile grammars, as well as dictionaries, that link graphemic, phonetic and semantic information for speakers and learners of the language to consult for text writing and speaking. The product approach to language forms also provides useful procedural tools for language and dialect comparison, dialect geography, language typology and historical linguistics.

But the situation changes when languages have been investigated for a very long time, such as English, German, French, Spanish, Arabic, Hindi, Japanese and Mandarin Chinese. When sound representations of words and structural representations of words in sentences have been put in systematic descriptive linguistic formats in such languages, linguistic pursuits may proceed in two different ways.

(1)   The formal representations may acquire a purpose in themselves and assume the status of the 'real' thing they are supposed to map. Then proponents of another linguistic paradigm may recycle the same data in a different format of their own, suggesting that it increases the explanatory power for the 'real' thing. So, we experience recycling of the same data from Structural Linguistics to Generative Grammar to Government and Binding, to Head-driven Phrase Structure Grammar to Role and Representation Grammar, and so on. An example from phonology is the treatment of Turkish vowel harmony in the frameworks of structural phonemics, generative phonology and Firthian prosodic analysis (Lees 1961; Voegelin and Ellinghausen 1943; Waterson 1956). The contribution of such *l'art pour l'art* linguistics to

the understanding of speech communication in human interaction is limited.

(2) On the other hand, it may be considered timely to renew theoretical reflection on how speakers and listeners interact with each other, using language beside other communicative means in contexts of situation. The forms obtained through a linguistic product approach will now be studied as manifestations of communicative functions in interactive language use. SFL, DRT and ISM are no longer discovery procedures, but theoretical models. They stop short, however, of reaching the dynamic level of speech interaction because they are still product-oriented and incorporate interaction context statically into structural representation.

Future research will benefit from advancing models of speech communication in interaction for at least some of the well-studied languages of the world. This monograph is an attempt in this direction, focusing primarily on German and English, but additionally including other languages in the discussion of selected communicative aspects. The results of this action approach can, in turn, be fed back into the product approach of language description and comparison. For example, in traditional language descriptions interrogative structures are compared between languages with regard to some vague 'question' concept. In the action approach, different types of question are postulated as different APPEAL functions in human interaction, and the *interrogative* forms found in different languages are related to these functions. This will have a great effect on making language teaching and language learning, based on linguistic descriptions of languages, more efficient.

Since, in addition to the syntactic structures, prosody is another central formal device in this functional framework, a prosodic model needs to be selected that guarantees observational and explanatory adequacy for the communicative perspective. This goal can best be achieved when the choice follows from a critical comparative overview of the most influential descriptive paradigms that have been proposed in the past. Therefore, the remaining section of this chapter provides such an historical survey to prepare the exposition, in Chapter 2, of the prosodic model adopted for integration in the *Organon Model*.

1.4   Descriptive Modelling of Prosody – An Overview of Paradigms

The study of prosody has concentrated on intonation and, with few exceptions, such as Bolinger's work (1978, 1986), has focused on the formal elements

and structures of auditory pitch of acoustic F0 patterns. Questions of meaning and the function of these patterns were raised *post hoc*, above all in relation to syntactic structures, sentence mode, phrasing and focus. Two influential paradigms in the study of prosody, the British and the American approach, are briefly discussed here, as a basis for the exposition of the Kiel Intonation Model (KIM), the former because KIM is an offspring of it, the latter in order to show and explain the divergence of KIM from present-day mainstream prosody research. Examples will, in each case, be presented in original notations, as well as in KIM/PROLAB symbolisations (cf. the list at the end of the Introduction), for cross-reference.

1.4.1   The Study of Intonation in the London School of Phonetics

Descriptions of intonation by the London School of Phonetics (Allen 1954; Armstrong and Ward 1931; Cruttenden 1974, 1986 (2nd edn 1997); Jones 1956; Kingdon 1958; Lee 1956; O'Connor and Arnold 1961; Palmer 1924; Palmer and Blandford 1939; Schubiger 1958; Wells 2006) relied on auditory observation and introspection for practical application in teaching English as a foreign language. Armstrong and Ward (1931) and Jones (1956) set up two basic tunes for English, imposed on stress patterns and represented by dots and dashes and curves: Tune I falling, associated with statements, commands and *wh* questions, Tune II rising, associated with requests and word-order questions. Modifications of these generate falling-rising and rising-falling, as well as pitch-expanded and compressed, patterns, signalling emphasis for contrast and intensity.

Palmer, Kingdon, and O'Connor and Arnold elaborated this basic two-tune concept by differentiating tunes according to falling, low-rising, high-rising, falling-rising and rising-falling patterns. Palmer introduced tonetic marks in orthographic text to represent the significant points of a tune, rather than marking every syllable. This was a move towards a phonological assessment of prosodic substance. The tune was also divided into syntagmatic constituents. O'Connor and Arnold's practical introduction became the standard textbook of Standard Southern British English intonation, proposing a division of tunes, now called tone groups, first into nucleus and prenucleus, then into nuclear tune and tail, and into head and prehead, respectively. These structural parts, with their paradigmatic elements, are combined into ten Tone Groups, five with falling, five with rising tunes at the nucleus. The intonation patterns are, in turn, related to four grammatical structures – statements, questions, commands and interjections. These are formal syntactic structures: declarative syntax, lexical interrogative (called special questions), word-order interrogative

syntax (called general questions), imperative syntax and interjectional ellipsis. High-rising nuclear tunes in declarative syntax ('You like him?') are discussed under the formal heading of statements, though referred to as 'questions' in a functional sense. Similarly, low-falling nuclear tunes in word-order question syntax ('Will you be quiet!', 'Stand still, will you!', with a high head, or 'Aren't you lucky!', with a low head) are discussed under the formal category of 'general questions', though referred to as 'commands' or 'exclamations' in a functional sense. This highlights the formal point of departure of intonation analysis. However, the formal description is followed by a discussion of fine shades of meaning carried by the ten tone groups in their four syntactic environments. This discussion is couched in descriptive ordinary-language word labels (e.g. 'Tone Group 2 is used to give a *categorical*, *considered*, *weighty*, *judicial*, *dispassionate* character to statements'), not in terms of a semantic theory of speech functions. The result is a mix of the formal elements and structures of intonation and syntax in English with *ad hoc* semantic interpretations. The descriptive semantic additions include attitudinal and expressive meaning over and above the meaning of syntax-dependent sentence modes, i.e. they are treated inside linguistics, not relegated to paralinguistics.

The phoneticians of the London School were excellent observers, with well-trained analytic ears. Although they did not have the concept of *alignment* of pitch accents with stressed syllables, and did not separate *edge tones* from *pitch accents*, central premises in AM Phonology, they described the auditory differences in minute, accurate detail. What AM Phonology later categorised as H+L*, H* or L+H*, L*+H pitch accents, combined with L-L% edge tones, are separate unitary pitch contours in the taxonomic system of the London School: low fall, high fall, rise-fall. AM/ToBI H* and L+H*/L*+H, combined with L-H%, are fall-rise and rise-fall-rise. Ladd (1996, p. 44f, 122f, 291 n.6, 132ff) accepts this contour approach as observationally adequate but does not consider it descriptively adequate, because it does not separate edge tones from pitch accents and does not associate the latter with stresses in various alignments.

In Ladd's view, a lack of insight into prosodic structures is most obvious in the way the London School phoneticians treat (rise-)fall-rises in British English. He argues that a rise-fall-rise pattern is compressed into a monosyllabic utterance, but is not spread out across syllables following a stressed syllable. In this case, the fall occurs on the nuclear syllable, the rise at the end of the utterance, with syllables on low pitch in between. To illustrate this he gives the example:

i.   A: I hear Sue's taking a course to become a driving instructor.

(a) B: Sue!? [L*HL-H%].
             PROLAB: **2**(Sue **&., &PG**

             A [L*+H] driving instructor [L-H%]!?
             PROLAB: A **2**(driving instructor **&., &PG**

The low tone of the combined pitch accent L*+H is associated with the stressed syllable of 'driving', the trailing high tone with this stressed and the following unstressed syllable. The low tone of the phrase accent L- is associated with the second syllable of 'driving' and the first two syllables of 'instructor', creating a long low stretch, and the high tone of the boundary tone H% is linked to the final syllable. This shows, according to Ladd, that the edge tones L-H% must be separated from the pitch accent in both cases, although they form an observable complex pitch contour on the monosyllable. The analysis with AM categories and ToBI symbols leaves out an important aspect of the actual realisation, which can be derived from this phonological representation in combination with the impressionistic pitch curve that Ladd provides. The final-syllable pitch rise after a stretch of low pitch gives the stressed syllable of 'instructor' extra prominence, partially accenting the word. The pitch pattern is thus turned into a rise-fall on main-accent 'driving', followed by a rise on partially accented 'instructor'.

i.   (b) B: *PROLAB:* A **2**(dr'iving **&2. &1**[instr'uctor **&, &PG**

This is no longer the same pattern as the rise-fall-rise on the monosyllabic utterance, and would not convey the same intended meaning. Therefore, Ladd's line of argument is no proof of a need to separate edge tones from pitch accents in intonational phonology.

The structurally adequate systematisation of rise-fall-rise intonations in English becomes a problem in Ladd's analysis, rather than in that of the English phoneticians, because, in the wake of AM Phonology, Ladd does not distinguish between unitary fall-rise and sequential fall+rise intonation patterns, which were separated as meaningful contrasts by the London School, especially by Sharp (1958). Prosodically the two patterns differ in the pitch end points of the fall and of the following rise, being lower for both in the sequence F+R than for the unitary FR, and they also differ in rhythmic prominence on the rise of F+R, as against FR, resulting in a partial accent on the word containing the rise. If the partial accent is put on a function word it naturally has a strong form, whereas in FR a weaker form occurs. This is an additional manifestation of greater prominence in the rise of F+R. Sharp provides an extensive list of

examples for both patterns, predominantly in statements and requests, a few in information and polarity questions, and some miscellaneous cases. He is less sure about the occurrence of FR in questions, but maintains, against Lee (1956, p. 70) and Palmer (1924, p. 82), who mention its absence from this sentence mode, that it does occur, but less frequently than in the other modes. It seems to be perfectly clear, however, 'that in both "yes-no" questions and "special" questions at least one focus for the patterns is quite common: the *first* word [of the question]. FR, in these circumstances, asks for confirmation or repetition, F+R pleads for an answer (or for action)' (Sharp 1958, p. 143). Sharp does not give any examples 'for these circumstances', but from the general functional description he has given for FR and F+T in questions, the following typical instances may be constructed:

ii.   (a)   [FR] What did you say?
                'I did not catch that, please repeat.'
                PROLAB: **&2^**What did you say **&., &PG**
                The fall before the rise adds insistence to the request for repetition, which is absent in a simple rise starting on 'what':
                PROLAB: **&2]**What did you say **&, &PG**

       (b)   [F] What did you [R] say?
                'Give me the content of what you said (when he asked you).'
                PROLAB: **&2^**What did you **&2. &1]**say **&, &PG**

       (c)   But a full accent on the rise is more likely:
                'Tell me what you said (when he asked you).'
                PROLAB: **&2^**What did you **&2. &2[**say **&, &PG**

           The fall before the rise in (b) and (c) adds insistence to the request for information, which is absent when the rise on 'say' is preceded by a high, instead of a falling, prenucleus:
                PROLAB: **&2^**What did you **&0. &2]** say **&, &PG**

iii.   (a)   [FR] Are you going to tell him?
                'He needs to be told, please confirm.'
                PROLAB: **&2^**Are you going to tell him **&., &PG**
                The fall before the rise adds insistence to the request for confirmation, which is absent in a simple rise starting on 'are':
                PROLAB: **&2]**Are you going to tell him **&, &PG**

       (b)   [F] Are you going to [R] tell him?
                'Inform me whether you will tell him.'
                PROLAB: **&2^**Are you going to **&2. &1]**tell him **&, &PG**

       (c)   But a full accent on the rise is more likely:
                PROLAB: **&2^**Are you going to **&2. &2[**tell him **&, &PG**

> The fall before the rise in (b) and (c) adds insistence to the request for information, which is absent when the rise on 'tell' is preceded by a high, instead of a falling, prenucleus:
> *PROLAB:* **&2^**Are you going to **&0. &2**]tell him **&, &PG**

In examples (ii.a) and (iii.a), the peak of FR has medial alignment with the accented syllable, AM H*, PROLAB **&2^**. In (ii.b) and (iii.b), a partial accent is possible for the rise of F+R on 'say' or 'tell him', but the full accent in (c) conveys the given meaning more clearly. The increased prominence that signals it is produced by the F0 onset of the rise in the accented syllable being critically below the end point of the preceding fall. This difference between a partially and a fully accented rise in F+R cannot be represented in the London School framework because accent is not a separate category from intonation and rhythmic structure. The examples in (ii.) and (iii.) have been constructed on the basis of Sharp's description. There are one or two examples in his list of the FR and F+R distinction in initial focus position of questions, but they are different from the ones in (ii.) and (iii.); they represent his standard patterns of medial-to-late FR alignment and F+R accentuation.

iv.　　(a)　[FR] What's his name?
　　　　　　　a 'I have forgotten.' *b* 'I am incredulous.'
　　　　　　　PROLAB: **&2^-(**What's his name **&., &PG**

　　　　(b)　[F] What shall I [R] tell him?
　　　　　　　'I really cannot think of anything.'
　　　　　　　PROLAB: **&2^**What shall I **&2. &1**]tell him **&, &PG**
　　　　　　　Accent **&2[**is possible as well when 'tell' is given a second major information point.

v.　　　　　[F] Are you [R] coming?
　　　　　　　'Do tell me whether you are coming.' 'Must I wait here for ever?'
　　　　　　　(Despair)
　　　　　　　PROLAB: **&2^**Are you **&2. &1**]coming **&, &PG**

Sharp did not distinguish clearly between two different alignments of FR. Except for the cases illustrated in (ii.) and (iii.), his examples refer to medial-to-late alignment of FR with the accented syllable. His FR data also appear to be all of the non-intensified type of (rise-)fall-rise, and therefore do not correspond to the AM category L*+HL-H% in Ladd's emphatic example, but to (L+)H*L-H% (*PROLAB:* **&2^**…**&.**, versus **&2^-(**…**&.**,). The general meanings of F+R and FR may be given as 'associative' versus 'dissociative' reference to alternatives in preceding speech actions. Here are two sets of examples:

vi. A: Look, there's Peter.
     B: I've seen him.
     (a) [aɪv FR siːn ɪm] 'I saw him before you even pointed him out.'
          PROLAB: [aɪv **2^-**(siːn ɪm **&., &PG**]
     (b) [aɪv F siːn R hɪm] 'I have spotted the person you are pointing to.'
          PROLAB: [aɪv **&2^**siːn **&2. &1**]hɪm **&, &PG**]
     (c) [aɪv FR siːn R hɪm] 'I saw the person you are pointing to without
          you mentioning it.'
          PROLAB: [aɪv **2^-**(siːn **&., &1**]hɪm **&, &PG**]

In FR of (a) 'him' has its weak form, in F+R of (b) its strong form. (c) shows that an FR on 'seen' may be followed by a simple rise on 'him' [hɪm] (again in its strong form, as in (b)), giving it more prominence, and partially accenting and foregrounding it. This rules out an association of the rise of the (rise-)fall-rise with an edge tone and is therefore outside the scope of AM Phonology.

There are further possibilities:

vi. (d) [aɪv F siːn ɪm], 'reporting the fact that I have seen him'
          PROLAB: [aɪv **&2^**siːn ɪm **&2. &PG**]

     (e) [aɪv F siːn hɪm]
          with partial accent on 'him', like (d) but foregrounding 'him'.
          PROLAB: [aɪv **&2^**siːn **&2. &1**)hɪm **&2. &PG**]

     (d) and (e) differ from (a) and (c) by only reporting speaker-oriented facts, whereas the latter involve the dialogue partner.

vii. A: You chaired the appointment committee for the chair of phonetics. The committee decided to take the applicant from down-under. Was it a good choice?
     B: I [F] thought [R] so. 'That was my opinion and it still is.'
          PROLAB: I **&2^**thought **&2. &1**]so **&, &PG**
          I [FR] thought so. 'That was my opinion at the time, but I have changed my mind.'
          PROLAB: I **&2^-**(thought so **&., &PG**

These data, analysed with observational as well as descriptive adequacy in the London School of Phonetics, cannot be handled in the AM Phonology framework, precisely because it links the rise to edge tones. Intermediate phrase boundaries cannot be introduced to solve the problem because there are no phonetic grounds for them. This had already been pointed out with reference to German data in Kohler (2006b, pp. 127ff), cf. 2.7. In addition to pitch accent L*+H, followed by the edge tones L-H%, Ladd (1996, p. 122) discusses some examples in British English for which he postulates pitch accent H*:

viii.   (a1)   Could I [H*] have the [H*] bill please [L-H%]?
                 PROLAB: Could I **&2^**have the **&0. &2^**bill please **&., &PG**

         (b1)   Is your [H*] mother there [L-H%]?
                 PROLAB: Is your **&2^**mother there **&., &PG**

They sound 'condescending or peremptory' to speakers of North American English, where a high-rising nucleus + edge tones, H*H-H%, would be used instead:

viii.   *(*a2)   Could I [H*] have the [H*] bill please [H-H%,]?
                 PROLAB: Could I **&2^**have the **&1. &2]**bill please **&? &PG**

         (b2)   Is your [H*] mother there [H-H%]?
                 PROLAB: Is your **&2]**mother there **&? &PG**

The reference to Halliday's broken Tone 2 in viii. (p. 291 n.6) makes it clear that Ladd is referring to a fall (not a rise-fall) on the accent of 'bill' or 'mother', followed by a rise on unaccented 'please' or 'there' in word-order questions. The pattern is a unitary *fall-rise*, making an *associative* reference to preceding actions of the type 'I've been served, I've eaten, I want to pay now' in (a), and 'I would like to speak to your mother. Is she in?' in (b). In both cases the rise establishes contact with the person spoken to; a simple fall would lack this and sound abrupt.

These examples could, of course, also be spoken with a unitary *rise-fall-rise*, and would then make *dissociative* references, (a) 'Waiter, I've been trying to catch your attention but you are constantly dealing with other customers, I am in a hurry' (b) 'Sorry, it's not you I have come to see, but your mother.' And in (a), 'please' may get extra prominence, giving it a secondary accent, in a separate rise after a fall or a fall-rise, creating F+R or FR+R and adding insistence to the request.

viii.   (a3)   *PROLAB:* Could I **&2^**have the **&1. &2^-(**bill please **&., &PG**
         (a4)   *PROLAB:* Could I **&2^**have the **&1. &2^**bill **&2. &1]**please **&, &PG**
         (a5)   *PROLAB:* Could I **&2^**have the **&1. &2^-(**bill **&., &1]**please **&, &PG**
         (b3)   *PROLAB:* Is your **&2^-(**mother there **&., &PG**

Parallel to the British English example (viii.b1) 'Is your mother there?', Ladd (1996, p. 122) discusses the German equivalent in the AM Phonology framework:

ix.   (a1)   Ist deine [H*] Mutter da [L-H%]?
                 probably based on an exponency classifiable as
                 Ist deine [FR] Mutter da?

and as *PROLAB:* Ist deine **&2^Mutter** da **&., &PG**
But there are other possible realisations.

(b1)   Ist deine [F] Mutter [R]da?
PROLAB: Ist deine **&2^Mutter &2. &1]**da **&, &PG**
partially foregrounding 'being *present*' as a minor information
point beside the main information point 'your mother'

(a2)   *PROLAB:* Ist deine **&2^-(Mutter** da **&., &PG**
(b2)   *PROLAB:* Ist deine **&2^-(Mutter &,. &1]**da **&, &PG**

The functional interpretations of these patterns are the same as in the English equivalents.

1.4.2   *Halliday's Intonational Phonology*

Halliday followed the tradition of the London School of Phonetics, but he incorporated the phonetic analysis of intonation in a phonological framework within his categories of a theory of grammar (Halliday 1961). In two complementary papers (1963a,b), which were republished in adapted and more widely distributed book form in 1967, he described intonation as a complex of three phonological systemic variables, *tonality*, *tonicity* and *tone*, interrelated with a fourth variable, *rhythm*. Tonality refers to the division of speech events into melodic units, *tone groups*. The tone group enters into a hierarchy of four phonological units together with, in descending order, the *rhythmic foot*, the *syllable* and the *phoneme*, each element of a higher-order unit consisting of one or more elements of the unit immediately below, without residue. The rhythmic feet in a tone group form a syntagmatic structure of an obligatory *tonic* preceded by an optional *pretonic*, each consisting of one or more feet. This structure is determined by the tonicity variable, which marks one foot in the foot sequence of a tone group as the *tonic foot*, by selecting one of a system of five tonal contrasts, the *tones* 1 fall, 2 high rise, 3 low rise, 4 fall-rise, 5 rise-fall. Feet following the tonic foot in the tonic of a tone group generally follow the pitch course set by the tone of the tonic. Besides these single tonics there are the *double tonics* 13 and 53, uniting tone 1 or 5 with tone 3 in two successive tonic feet of the tonic section of one tone group. They form major and minor information points and correspond to F+R versus FR in tone 4.

Tied to the tone selection at the tonic there are further tone selections at the pretonic. At both elements of tone group structure, a principle of delicacy determines finer specifications, such as different extensions of the fall in tone 1 (1+ high, 1 mid, 1- low), different high-rising patterns for tone 2 (2 simple rise, 2 rise preceded by high fall: broken tone 2), and different extensions of the fall in tone 4 (4 mid fall-rise, 4 low fall-rise). Each rhythmic foot has a syntagmatic structure of obligatory *ictus*, followed by optional *remiss*; the former is filled

by a strong syllable, the latter by one or more weak syllables. Halliday follows Abercrombie (1964) in assuming stress-timed isochronicity for English, and that the ictus may be silent ('silent stress') 'if the foot follows a pause or has initial position in the tone group' (Halliday 1963a, p. 6).

Halliday integrates his intonational phonology into the grammar of spoken English, where the intonational systems operate side by side with non-intonational ones in morphology and syntax, at many different places in the coding of meaningful grammatical contrasts. In the 1963a paper, he looked from phonological contrasts to distinctive grammatical sets, asking 'What are the resources of intonation that expound grammatical meaning?', whereas in the 1963b paper, he looked at the phonological contrasts from the grammatical end, asking 'What are the grammatical systems that are expounded by intonation?' With this approach, Halliday took a step towards a functional view of phonological and grammatical form, which he has been concerned with ever since in the development of a coherent framework of *Systemic Functional Linguistics* (SFL).

Pheby (1975) and Kohler (1977 (1st edn)) applied Halliday's framework to German. They were an advance on von Essen (1964), who delimited three basic pitch patterns with reference to vaguely defined functional terms – terminal, continuative, interrogative intonation – and was then forced to state that *yes-no* questions have rising intonation, question-word questions and statements terminal intonation, and syntactically unfinished sentences continuation rises. This analysis, quite apart from being superficial and incomplete, mixed up the formal and functional levels of intonation right from the start, which the British colleagues and Kohler (1977, 1995, 2004, 2013b) did not; they knew, and said so, that both question forms can have either terminal or rising pitch with finer shades of meaning.

The more recent publication by Halliday and Greaves (2008) expounds the Hallidayan intonation framework in greater detail and reflects its integration with grammar in the very title. Whereas the earlier publications described the intonation of Standard Southern British English (RP), the later one includes Australian and Canadian English, thus taking 'English' in a more global sense, and it illustrates the descriptions with Praat graphics in the text and with sound files of isolated but grammatically contextualised utterances, as well as of dialogues, on an accompanying CDROM. Meaning as carried by intonation is now related to three of Halliday's four metafunctions: the interpersonal, the textual and the logical. The systems of *tonality* and *tonicity* are linked to textual meanings, the systems of *tone* to interpersonal meanings. The phonological rank scale is paralleled by a grammatical rank scale of *sentence*, *clause*,

group/phrase, *word*, *morpheme*, linking to experiential, interpersonal and textual meanings. Setting up separate systems for intonation and grammatical structure is a good principle because it avoids the conflation of falling or rising tonal movement with declarative and two types of interrogative structure, as has been quite common. But cutting across this grammatical rank scale is the *information unit*, which is not independently defined, and seems to be in a circular-argument relationship with the phonological unit of the tone group, since by default one tone group is mapped onto one information unit: 'Thus the two units, the phonological "tone unit" and the grammatical "information unit" correspond one to one; but since they are located on different strata, their boundaries do not correspond exactly. In fact, both are fuzzy: the boundaries are not clearly defined in either case' (Halliday and Greaves (2008), p. 99). This means that adding yet another unit to the extremely complex taxonomic intonation-grammar system does not seem to serve a useful purpose, and Crystal (1969b) had already criticised the concept in his review of Halliday (1967).

Another weak point of Halliday's intonational phonology concerns the division of the stream of sound into tone groups and of these into rhythmic feet. Although Halliday and Greaves gave up the doubtful isochrony principle and no longer quote Abercrombie (1964), rhythmic regularity is still the building principle of the tone group: 'When you listen carefully to continuously flowing English speech, you find there is a tendency for salient syllables to occur at fairly regular intervals, and this affects the syllables in between: the more of them there are, the more they will be squashed together to maintain the tempo' (Halliday and Greaves (2008), p. 55). This can be a useful heuristics when dealing with isolated sentences in foreign language teaching, even more so for learners whose native languages have totally different rhythmic structures from English, such as French. Teaching English as a Foreign Language was a prominent field of application of a large part of intonation analysis in the London School of Phonetics. Halliday, likewise, worked out his system of intonational phonology for the *Edinburgh Course in Spoken English* (1961) by R. Mackin, M. A. K. Halliday, K. Albrow and J. McH. Sinclair, later published by Oxford University Press (see Halliday 1970). The *Intonation Exercises* of this course were reproduced as teaching materials at the Edinburgh Phonetics Department Summer Vacation Course on the Phonetics of English for foreign students. In 1965 and 1966, I was asked to give these intonation tutorials.

But the rhythmic foot analysis of the tone group does not really provide a good basis for analysing continuous speech. Moreover, Halliday's intonational phonology lacks the category of a phrase boundary. Such a prosodic phrase

marker encapsulates a bundle of pitch, duration, energy and phonation features to signal a break, which may, but need not, coincide with grammatical boundaries and with the boundaries Halliday sets up for his tone groups. In sequences of rhythmic feet, Halliday earmarks those that contain one of his five tones, the tonic feet, constituting the tonics of tone groups. Since by arbitrary definition any one tone group can only have one tonic (except for the major+minor tonic compounds 13 and 53), there must be a tone group boundary between two succeeding tonics. Where this boundary is put is again arbitrary in view of the fuzziness Halliday and Greaves refer to in the quotation above, i.e. due to the lack of a phonetic criterion that determines a phrase boundary. This was again pointed out by Crystal (1969b). In many cases, Halliday no doubt takes the grammatical structure into account when deciding on the positions of tone group boundaries. But this is against his principle of setting up separate phonological- and grammatical-rank scales and relating them afterwards, and the violation of this principle borders on circularity.

And, finally, giving tone groups a rhythmic foot structure conflates rhythmic grouping into ictus and remiss with meaning-related phrasal accentuation. Halliday's framework does not provide a separate accent category outside the tonic, and in the latter it is the pitch-related tone category that determines the tonic foot and the tonic syllable, and thus constitutes a phrasal accent. The syllable string preceding the tonic may contain meaning-related phrasal accents, but not all ictus syllables of a postulated rhythmic foot structure are accented. A tonic foot may be preceded by a multisyllable *prehead*, which contains no accent, but may be perceived as a sequence of strong and weak syllables due to timing and vowel quality, for example before a tonic containing tone 3 in:

> // 3 don't stay / out too */ **long //** (Halliday and Greaves (2008), p. 119; see Figure 1.2a)

In Hallidayan notation 'don't' and 'out' are treated as ictus syllables in two rhythmic feet of the pretonic and a tone-3 tonic. But, when listening to the .wav file (supplied on the CDROM), no accent can be detected in the pretonic syllable sequence, and the perception of rhythmic structure fluctuates between the one noted and /don't stay out too/. The vocalic elements in all four syllables have durations between 120 and 130 ms. Duration would be considerably longer in an accented syllable containing a diphthongal element.

What the (male) speaker realises here is a high prehead before the (only) sentence accent, in a *high register* at a pitch level around 180 Hz, which at the same time increases the *pitch range* down to the following low rise. The speaker could, of course, have used a high prehead without going into a high register
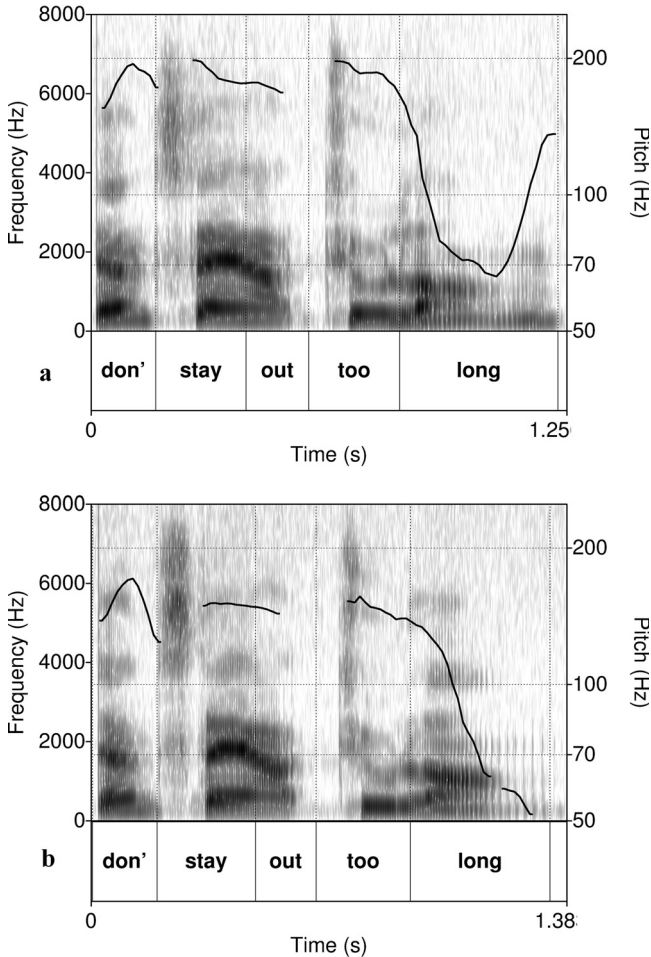
Figure 1.2. Spectrograms and F0 traces (log scale) of *a* // 3 don't stay / out too */ long // – audio file 5_2_2_4a3.wav, and *b* // 1 don't stay / out too */ long // – audio file 5_2_2_4a4.wav, from Halliday and Greaves (2008), p. 119. Standard Southern British English, male speaker (M. A. K. Halliday)

and thus without increasing the pitch range. In the high prehead, F0 fluctuation is largely conditioned by vowel-intrinsic and consonant-vowel coarticulatory microprosody: only the initial 'don't' has a more extensive rise, which, just like vowel duration, is not large enough to signal a phrasal accent. The listener may then structure the prehead rhythmically in variable ways. Halliday differs from the London School of Phonetics, e.g. O'Connor and Arnold (1961), by

not having the category of prehead. The composition of tone groups by rhythmic feet with an obligatory ictus syllable that may be silent precludes it.

How serious this omission is in a systemic functional approach to intonation is shown by the example:

> // 1 don't stay / out too */ **long //** (Halliday and Greaves (2008), p. 119; see Figure 1.2b)

The notation given for this tone group differs from the previous one only by having tone 1 instead of tone 3. But listening to the .wav file reveals two differences: (1) 'don't' is accented because its prominence is greater, due to longer duration of its sonorous part, and to more extensive F0 movement, well above the pitch level of the following syllables, so the pretonic sequence is not a prehead; (2) the pretonic sequence is at a much lower pitch level of 150 Hz – even the peak in 'don't' only reaches 170 Hz. The accent on 'don't', combined with the lower pitch level preceding the final fall, intensifies the meaning of a command, whereas the unaccented, but high prehead preceding the final low rise intensifies the meaning of a request, and the high register adds a note of entreaty. These are important aspects of the transmitted meanings, which are not reflected by different tonal categorisation in Halliday's notation: the two pretonics are identical because they are given the same rhythmic structure. But this rhythmic structure is an additional overlay on accentuation, register and range, and may surface perceptually in variable ways in both utterances. In PROLAB, the two utterances are differentiated as:

> **&HP &HR** don't stay out too **&2[**long **&, &PG**
> **&2^**don't stay out too **&0. &2^**long **&2. &PG**

The additional rhythmic structure is captured at the level of segmental spectrum and timing.

The following postulates of Halliday's intonational phonology can be taken as essential for any prosodic framework:

- English intonation is based on a system of contour-defined contrastive tones.
- Parallel to the phonological tone system there are lexicogrammatical systems.
- Phonological form is part of the grammar as another exponent of meaning in language functions.

But to be applicable to the analysis and description of prosodic systems in connected speech, more particularly spontaneous speech, and in text-to-speech

synthesis, several weak points of Halliday's systemic functional approach need adjusting.

- The nesting rank scale of phonological units, as well as the immediate constituents division of tone groups into tonic and pretonic, do not provide an adequate representation of prosodic structures – especially, the composition of the unit of the tone group by elements of the unit of the rhythmic foot cannot cope with the dynamic flow of speech and rhythmic disturbances such as hesitations, false starts, repetitions. Instead we need an accent category with several levels, based on degrees of prominence, to which tones are linked. In between successive accents, pitch is organised into distinctive concatenation patterns.
- Speech is organised into prosodic phrases, so prosodic phrase boundaries need to be determined by bundles of phonetic features.
- In prosodic phrases, the first accent may be preceded by unaccented preheads, and they form a system of mean, low and high pitch.
- Register needs to be introduced to set the pitch level of prosodic phrases, or of the part up to the final accent-linked pitch turn (thus also determining pitch range), or of sequences of prosodic phrases.

When these weaknesses of Halliday's intonational phonology became relevant in the Kiel TTS development (Kohler 1991a,b) and in spontaneous speech annotation for the Verbmobil project (Kohler, Pätzold and Simpson 1995), the description of German intonation, given in Hallidayan terms in the first edition of Kohler (1977), was put on a new basis developed for the tasks: the *Kiel Intonation Model*. It was presented in Kohler (1991a,b), then in the second edition of Kohler (1995) and in Kohler (1997a,b), and will be set out in Chapter 2. Subsequent chapters will take Halliday's form and function perspective one step further. Whereas Halliday looked from phonology to grammar and from grammar to phonology in the early papers, and later related phonological form in grammar to metalinguistic functions, I shall reverse the relationship, set up a few basic communicative functions within Bühler's model and then investigate language-specific prosodic, syntactic and lexical carriers for them.

1.4.3   Pike's Level Analysis

Pike laid the foundation for the analysis of American English intonation on a different descriptive basis, auditorily referring significant points of pitch contours – starting and ending points, and points of direction changes, in relation to stressed syllables – to four pitch levels, 1–4 from highest to lowest. Not every unstressed

syllable gets a significant pitch point but may have its pitch interpolated between neighbouring pitch points. On the other hand, a syllable may get more than one significant point to represent the stress-related pitch contour, or even more than two, when a contour changes direction and is compressed into a single stressed syllable. Pike gives a detailed formal account of the resulting pitch-level contours of American English and relates them to syntactic structures. He points out that the contours found in statements can also occur in questions and vice versa, and he provides a wealth of *ad hoc* references to attitudinal and expressive shades of meaning added to utterances by pitch contours. His analysis thus parallels the one by O'Connor and Arnold with a different paradigm for a different variety of English.

1.4.4   Intonation in AM Phonology and ToBI

As Halliday provided a phonological framework within structuralist grammar for the intonation analysis of the London School of Phonetics, Pierrehumbert put the Pikean level analysis of intonation into a framework of Autosegmental Metrical (AM) Phonology. The distinctive pitch levels were reduced to two, H and L, which, on their own and in the sequence H+L and L+H, form systems of pitch accents, phrase accents and boundary tones. In pitch accents, H and L are associated with stressed syllables indicated by *, but they may have leading or trailing H or L, yielding H*, L*, H+L*, H*+L, L+H*, L*+H. The separation of H* and L+H* was a problematic alignment category in AM Phonology and ToBI because a dip between two H* accents requires an L tone attached to an H* tone, given the principle of linear phonetic interpolation between pitch accents.

Falling, rising or (rising-)falling-rising nuclear pitch contours (of the London School), which in the extreme case are compressed into a one-syllable utterance, such as 'yes', are decomposed into three elements: a pitch accent, followed by a phrase accent and, then, by a boundary tone, in each case with selection of H or L. All three must always be represented, e.g. H*L-L%, H*H-H%, L*H-H%, H*L-H%, L+H*L-H%, L*+HL-H%. Since falling-rising contours are defined by three pitch points, three types of syntagmatic element are needed to represent them. AM Phonology selects them from the three accent and boundary categories and extrapolates them to all contours, including monotonic falls and rises. These phonological elements are associated with syllables and phrase boundaries, linked to F0 traces and aligned with segmental syllable structure in spectrograms. The confounding of pitch accents with edge tones has already been reviewed in the discussion of AM solutions for FR and F+R patterns of the London School in 1.4.1.

AM Phonology is a highly sophisticated formal framework, which, beyond the basic premises sketched here, has been undergoing continual change over the years, and right from the outset the focus has been on form, not on function and meaning. When the AM phonological framework became the basis for a transcription system, ToBI, the original strict language-dependent systemic approach began to get lost, phonetic measurement was squeezed into the preset categories, which were transposed to other languages, and the transcription tool was elevated to the status of a model.

Questions of meaning of the formal intonation structures have been raised, but *post festum*, for example by Pierrehumbert and Hirschberg (1990), who propose a compositional theory of intonational meaning related to pitch accents, phrase accents and boundary tones. Another, very influential example of linking intonational form to meaning is Ward and Hirschberg (1985), where the rise-fall-rise contour, based on the AM representation L*+HL-H%, is analysed as a context-independent contribution to conveying speaker *uncertainty*. It appears, however, that most of the examples discussed by Ward and Hirschberg are not instances of L*+HL-H%, but of L+H*L-H%, which they explicitly exclude as the phonological representation of their rise-fall-rise. With the L*+HL-H% pattern, a speaker is said to relate an utterance element to a scale of alternative values and to indicate not being certain whether the hearer can accept the allocation as valid. For example, in:

	B:	I'm so excited. My girlfriend is coming to visit tonight.
	A:	From far afield?
a.	B:	From suburban Phila\del/phia.
b.	B:	*From next \door/. (p. 766)

'[T]he speaker, a West Philadelphia resident, conveys uncertainty about whether, on a distance scale, *suburban Philadelphia* is *far away* from the speaker's location. … b. is distinctly odd, given the implausibility of B's uncertainty whether *next door* is *far away*' (p. 766).

The authors provide an analysis in terms of logical semantics at the Representation level, which considerably narrows the field of speech communication, and may thus make it difficult to capture the full range of the communicative function of the fall-rise pattern in English. If, in the above example, B were to give a facetious answer, with a smile on his face, b. would not be odd at all, but would be understood as an ironic reply to A's enquiry about distance. It would still be an instance of what Sharp (1958) called the dissociative reference to alternatives in his fall-rise FR. The semantic-prosodic distinction between this pattern and Sharp's F+R is nicely illustrated by the two versions

of the sentence 'I thought so' discussed in 1.4.1. The speaker expresses association with, or dissociation from, the earlier belief, using either F+R or FR, and is certain about that in both cases. With FR, the speaker is, on the one hand, definite about having changed his mind, by using a peak pattern, but on the other hand, plays it down in social interaction conforming to a behavioural code, by adding a rise to alleviate the categoricalness in an appeal to the listener to accept the change of mind. If the speaker makes a statement about the present opinion without associative or dissociative reference to the past, it may be 'I [F/R] think so', with either a fall or a low rise for a definite or a non-committal response.

Whereas all the Ward and Hirschberg examples of American English have their fall-rise equivalents in Standard Southern British English, this may not hold for transposing Sharp's British English examples to American English. If the pattern distinctions do apply to both varieties, the conflation of pitch accents with edge tones and the lack of an accent category, separate from pitch, preclude the distinctive representations of the semantic-prosodic subtleties related to fall-rise pitch patterns. This may be illustrated by the following contextualisations:

> To provide sufficient seating at a family get-together, father A says to his two boys B and C

A We need more chairs in the sitting-room. Go and get two from the kitchen and a couple more from the dining-room.

B [Goes to the kitchen, comes back with two chairs, says to A]

  (a) There's [FR] another one in the kitchen.
        PROLAB: There's **&2^**another one in the kitchen **&., &PG**

  (b) There's [F+R] another one in the kitchen.
        PROLAB: There's **&2^**another one in the **&2. &1]**kitchen **&, &PG**

  (c) There's [FR] another one in the [R] kitchen.
        PROLAB: There's **&2^**another one in the **&., &1]** kitchen **&, &PG**

C [Goes to the dining-room, gets two chairs, comes back via the kitchen, says to A]

  (d) There's [F] another one in the [R] kitchen.
        PROLAB: There's **&2^**another one in the **&2. &2[**kitchen **&, &PG**

  (e) There's [FR] another one in the [FR] kitchen.
        PROLAB: There's **&2^**another one in the **&., &2^**kitchen **&., &PG**

In (a), B uses a rise-fall-rise that falls sharply to a low level on 'another', and then immediately rises again to mid-level at the end of 'kitchen', which is unaccented because it is integrated in a monotonic rise from 'one' onwards. This is Sharp's unitary FR, Halliday's tone 4, and L+H*L-H% in AM Phonology. B transmits the meaning 'There's an additional chair in the kitchen, besides the two I have just brought from there, although Dad thought there were only two', a dissociative reference to alternatives.

In (b), the rise after the low-level fall on 'another' is delayed until 'kitchen', which is partially foregrounded with a partial accent. This is Sharp's compound F+R, and Halliday's double-tonic with tone 13. However, the pattern cannot be represented in AM Phonology because the categorisation L+H* L*L-H% for a fall followed by a rise, with two pitch accents and final edge tones, allocates two full accents to the phrase, and therefore does not distinguish (b) from (d). The F+R pattern makes an associative reference to alternatives; it does not have the contrastive reference to A's mention of 'two chairs from the kitchen'.

In (c), B makes the same dissociative reference to alternatives as in (a) but partially foregrounds 'kitchen', giving it a partial accent by breaking the rising contour of the fall-rise and by starting another rise from a lower level within the same intonation phrase. In Sharp's analysis, 'another' would receive a fall-rise FR, 'kitchen' a simple rise. Similarly, Halliday would have tone 4 followed by tone 3 in two tone groups. AM Phonology cannot represent this pattern because an intermediate intonation phrase would have to be postulated even in the absence of any phonetic boundary marker. If the pitch break were to be taken as the indication of such a phrase boundary, from which the presence of edge tones would in turn be deduced, the argument becomes circular. In all three descriptive frames, the different accent level of 'kitchen' versus that of 'another' would not be marked, and therefore the different meaning from (d) and (e) could not be captured.

Since C has brought chairs from the dining-room he refers contrastively to an additional chair in the kitchen, and gives 'kitchen' a full accent. In (d), Sharp's F+R is separated into F and R linked to the two accents, with associative reference to alternatives. Halliday would have to have two tone groups //1 There's another one //3 in the kitchen.// This analysis is independent of the presence or absence of phonetic boundary markers. In AM Phonology, the pattern may be represented by two pitch accents in one intonation phrase, L+H* L*L-H%, because the L of the second pitch accent provides the right-hand pitch point for linear interpolation of the fall from the H of the first pitch accent. In (e), there are dissociative references to an alternative number

of chairs and to an alternative locality, by two rise-fall-rises linked to the two accents. As in (c), there may again be a single prosodic phrase. Sharp's analysis would simply have FR in both positions; Halliday would again need two tone groups, each with tone 4. In AM Phonology, two intonation phrases with L+H*L-H% would be necessary to generate the four-point rise-fall-rise contours, each with two intonation-phrase edge tones in addition to two pitch-accent tones, irrespective of the potential absence of phonetic boundary markers between them.

Thus, the AM phonological representations in (d) and (e) of C differ in the relative allocation of prosodic information to the theoretical categories of paradigmatic *pitch accent* and syntagmatic *intonation phrasing*. This different allocation is conditioned by constraints in the canonical AM definitions of prosodic categories:

- Pitch accents are defined by up to two sequential H or L tones.
- In a sequence of pitch accents, the pitch contour between abutting tones is the result of linear phonetic interpolation between the phonological pitch-accent tones. Therefore, for example, in two successive peak patterns, a distinctive pitch dip between two H* necessitates postulating a bitonal pitch accent, either a trailing L tone in the first, or a leading L tone in the second.
- The pitch contour between the last pitch-accent tone and the end of the intonation phrase is represented by two sequential H or L edge tones, a phrase accent and a boundary tone.
- A rise-fall-rise intonation contour around an accented syllable, with four distinctive pitch points, must be represented by a bitonal pitch accent followed by two edge tones.
- If a rise-fall-rise contour occurs utterance-internal, it must be followed by an intonation phrase boundary.
- If there are no phonetic boundary markers indicating such a boundary, such as segmental lengthening, with or without a following pause, there are no pitch-independent reasons for postulating such a boundary, or the argumentation becomes circular by using pitch as the defining feature for the postulated boundary, which in turn determines the edge tones before it.

These constraints on the phonological representation of intonation contours in AM Phonology reduce descriptive and explanatory adequacy in prosodic data interpretation, compared with the accounts provided by the London School and Halliday.

1.4.4.1 Alignment of Rise-Fall-Rises in English

AM Phonology conceptualises English L*+HL-H% and L+H*L-H% rise-fall-rise patterns as different alignments of the L and H tones of the rise-fall pitch accent with the stressed syllable: either L or H is aligned with it, H trailing L* and L leading H*, producing later (delayed) or earlier association of the pitch accent with the stressed syllable. An even earlier alignment is given as H*L-H%, and there is a fourth possibility – H+L*L-H%, where alignment occurs with the syllable preceding the stressed one, which appears not to be discussed in the AM literature. KIM treats these pitch patterns as distinctive points on a scale of synchronisation, from *early* to *late*, of F0 peak maximum with vocal-tract timing, and uses the PROLAB notations **<&2) &.,>**, **<&2^ &.,>**, **<&2^-(&.,>**, **<&2(&.,>** (see 2.7).

Pierrehumbert and Steele (1987, 1989) raised the question as to whether the L+H*L-H% versus L*+HL-H% distinction is discrete or scalar. They based their investigation on the utterance 'Only a millionaire', with initial stress on the noun and F0 peaking earlier or later in relation to the offset of /m/. They contextualised the two versions in a scenario of a fund-raising campaign targeting the richest. A potential donor, when approached as a billionaire in a telephone call, replies, 'Oh, no. Only a millionaire', with L+H*L-H%, whereupon the charity representative expresses his incredulity and uncertainty with the later peak alignment L*+HL-H%. To decide on the discrete versus scalar issue, the authors performed a perception-production experiment. They took a natural production of a L+H*L-H% utterance as the point of departure for LPC synthesis, shifting the stylised rise-fall pattern in 20 ms steps through the utterance, with peak positions ranging from 35 ms to 315 ms after /m/ offset.

Five subjects were asked to listen to each of the fifteen stimuli in fifteen randomised blocks, and to imitate what they had heard. These imitations were recorded and analysed with the hypothesis that, if the categories are discrete, the ideal speaker/listener will allocate the percepts to two different categories and then reproduce them in such a way that the realisations will show a bimodal clustering. The statistical basis of this experiment is weak, not only because of the insufficient number of subjects, but more particularly since one hearer-speaker was the junior author, who, of course, knew what the test categories were and sounded like, and who produced the clearest bimodal pattern. Furthermore, one subject failed to produce even a vague resemblance of bimodality.

The authors' conclusion that the two phonological categorisations of rise-fall-rise patterns in AM Phonology represent a discrete contrast can

therefore not be accepted as having been proved. It is to be assumed that peak shifts in rise-fall-rise patterns are perceptually processed in similar ways to *peak* shifts in rise-fall patterns, as obtained for English and German (see 2.8). These data show that the perception of peak synchronisation only changes categorically from *early* (pre-accent) to *medial* (in-accent) position, but not for peak shift inside the accented vowel, from *medial* to *late*, where changes are perceived along a continuum. Since the Pierrehumbert and Steele experiment only dealt with the in-accent shift, the potential categorical change from pre-accent to in-accent could not become a research issue, and in view of the weakness of the experimental paradigm, the results do not support discrete patterning. The perceptual and cognitive processing of rise-fall-rise peak shifts may be considered parallel to that observed for rise-falls, with the addition of an interactional rapport feature carried by final rising pitch. Whereas in a shift from *early* to *medial* peak there is a discrete semantic change from Finality to Openness, coupled with a categorical perceptual change (see 2.8), the shift from *medial* to *late peak* successively adds degrees of Contrast and of the expression of Unexpectedness along a continuum of peak synchronisation. Furthermore, this expression includes other prosodic variables besides F0 alignment, i.e. F0 peak height, timing, energy and more breathy phonation.

This issue was investigated by Hirschberg and Ward (1992). They report recording the pattern L+H*L-H% with eight utterances in an 'uncertainty' as well as in an 'incredulity' context, where the latter was hypothesised to generate an expanded pitch range, different timing, amplitude and spectral characteristics. The utterances differed widely in the stretch of speech over which the rise-fall-rise was spread, with 'ELEVEN in the morning' at one end of the scale and 'Nine MILLION' at the other. For the former, the two contexts, as well as the F0 displays of the two data samples produced, are provided:

'uncertainty'	A	So, do you tend to come in pretty late then?
	B	\ELEVEN in the morning/.
'incredulity'	A	I'd like you here tomorrow morning at eleven.
	B	!ELEVEN in the morning!

! ! is to symbolise the incredulity version of the utterance with the same pitch-accent and edge-tone pattern L*+HL-H% as in the uncertainty version \/. The two figures provided show that F0 sets in low and starts rising at the end of the stressed vowel of 'eleven', peaks at the end of the accented word, stays high during the following vowel and then descends to a low level in 'the'. There

follows a further small F0 drop in the stressed vowel of 'morning', before F0 rises again in the final syllable. The two displays differ only in the F0 range, which is wider in the 'incredulity' version, with a slightly higher precursor and considerably higher peak and end points. These F0 patterns suggest that 'morning' received extra prominence and was accented in both cases. This would also be a more plausible realisation of the utterance in the two contexts than the one with a single accent on 'eleven' and a much earlier rise, starting somewhere around 'the'. So, this pattern looks different from a single-accent rise-fall-rise in 'million' and does not seem to be L*+HL-H%, but L*+HL*L-H%, a fall followed by a rise, as in (b) or (d) of 'There's another one in the kitchen' in 1.4.4. This would mean that 'incredulity' is signalled by the expanded pitch ranges of the *late peak*, which signals expressively evaluated CONTRAST, and of the *final rise*, probably supported by non-modal phonation. A double-accent fall-rise in the 'uncertainty' context does not make a dissociative reference to other alternatives, as the single-accent rise-fall-rise would. But the *late peak* contrasts, and expressively evaluates, B's time reference with A's question about coming in 'pretty late', and the *final rise* establishes contact with the dialogue partner and alleviates the categoricalness of a *late peak*.

Hirschberg and Ward used the recordings of the eight contextualised utterances to generate two sets of stimuli, categorised as conveying 'uncertainty' and 'incredulity' for a listening experiment, where subjects had to allocate each stimulus to one of the two categories. Since the pitch patterns were most probably not homogeneous, and since such context-free semantic allocations are difficult, especially in view of the somewhat opaque meaning of 'uncertainty', the conclusions about the physical properties that cue 'uncertainty' or 'incredulity' are not so clear as they are made out to be.

1.4.5 A New Paradigm

The critical historical survey in 1.4 has prepared the ground, and provided the rationale, for presenting a new paradigm. The following chapters model prosody in relation to communicative functions of speech interaction, on the basis of the Kiel Intonation Model (KIM) in a broad linguistic-paralinguistic setting. The concern for function in prosody research at Kiel University goes back to Bill Barry's paper 'Prosodic functions revisited again!' (Barry 1981), following Brazil (1975, 1978). The function perspective guided the analysis, in production and perception, of prosody in general, and of intonation in particular, from the early 1980s onwards, converging on the development of a prosodic model (Kohler 1991a,b, 1997b, 2006b, 2009b).

The idea behind KIM is that modelling prosody should mirror its use by speakers and listeners in communicative action, i.e. prosodic categories must be an integral part of communication processes rather than just static elements in a linguistic description. Speakers use prosody to structure the flow of sound for the transmission of meaning to listeners. In a synsemantic field, prosody operates on linguistic signs in parallel to morphological and syntactic patterning for propositional representation, and in a sympractical deictic field it signals Speaker-Listener-Situation relations. Finally, speakers use prosody to express their emotions and attitudes, and to signal their appeals to listeners. The prosodic model is to be structured in such a way that it can capture and adequately represent all these communicative functions in speaker–listener interaction. This also implies that the model needs to be integrated into a theory of speech and language together with all the other formal means – segmental-phonetic, lexical, morphological, syntactic – contributing in varying proportions as carriers of these functions. The model must be oriented towards basic communicative functions of *homo loquens*, and at the same time it must take into account psycho-physical components of the human-speech producing, perceiving and understanding mechanisms, irrespective of any particular language form that organises the general psycho-physical prerequisites for communicative purposes in language-specific ways.

KIM follows the European tradition of postulating a system of distinctive global pitch contours – *peak*, *valley*, combined *peak-valley* and *level* patterns. The model sets out how these patterns are synchronised with vocal-tract articulation, how they are concatenated into a hierarchy of larger units from phrase to utterance to paragraph in reading or to turn in dialogue, and how they are embedded in other prosodic patterns – vocal-tract dynamics, prominence and phonation, paying attention to both the production and the perception of prosody in communicative function. The model was developed over many years, starting with a project in the German Research Council programme 'Forms and Functions of Intonation' in the 1980s (Kohler 1991c), continuing with its implementation in the INFOVOX TTS system (Kohler 1997a) and with the development of a data acquisition and annotation platform in the PHONDAT and VERBMOBIL projects of the German Ministry of Research and Technology (Kohler, Pätzold and Simpson 1995; Scheffers and Rettstadt 1997). In this research environment, large databases of read and spontaneous German speech were collected (IPDS 1994–2006; Kohler, Peters and Scheffers 2017a–b) and annotated segmentally and prosodically with the help of the PRO[sodic]LAB[elling] tool (Kohler 1997b; Kohler, Peters and Scheffers 2017a–b),

which was devised to symbolise the prosodic systems and structures of KIM for computer processing of the German corpora. In a subsequent German Research Council project, 'Sound Patterns of German Spontaneous Speech', various prosodic aspects of the corpus data were analysed in the KIM-PROLAB frame (Kohler, Kleber and Peters 2005). PhD theses by Benno Peters (2006) and Oliver Niebuhr (2007b) followed, and there has been a continuous flow of prosodic research within this paradigm in Kiel.