Developing a Framework of Communicative Functions for the Study of Speech Prosody

Klaus J. Kohler

University of Kiel, Kiel, Germany

Abstract

This is a proposal in favour of proceeding from communicative function to linguistic form, rather than the reverse, for an insightful account of how humans communicate by speech in languages. A functional framework is developed that encompasses argumentation structures, declarative and interrogative functions, and expressive intensification. Such a function orientation can become a powerful tool in comparative prosodic research across the world’s languages. The potential of this approach is shown by comparing the prosodic form of Mandarin Chinese data collected in functionally contextualized scenarios with corresponding data from English.

Index Terms: linguistic form, communicative function, argumentation structure, declarative, interrogative, emphasis

1. Prosody in Structural Linguistics

The core of mainstream structural linguistics is linguistic form, abstracted from phonetic substance, and represented by discrete systemic units and their structural relations in syntax, morphology and phonology for the differentiation of meaning in the grammars of languages. In this paradigm, discrete formal contrasts constitute the pivot to which both phonetic manifestation and semantic interpretation, restricted to propositional meaning, are referred post hoc. Therefore graded differences between speech elements, as well as attitudinal and expressive meaning, are relegated to paralinguistics, i.e. outside linguistics proper, and communicative functions in speech interaction are conceptualized as being subsidiary to linguistic form. This also excludes a direct link between communicative function and the physics of speech behaviour. These restrictions of formal linguistics have had far-reaching effects on the shaping of research into prosody, which has been incorporated in formal phonology in the course of the past three decades.

Prosody in speech interaction is the phonetic manifestation par excellence to jointly code information-related (propositional), listener-related (appellative), and speaker-related (expressive) meaning at any given moment, in different proportions, depending on the communicative function in the particular situational context [1]. Therefore, a model of prosody that does not incorporate a communicative framework in its own right beside the formal level, to relate linguistic form and phonetic substance to communicative functions in all three semantic fields, is not capable of providing a comprehensive, descriptively and explanatorily adequate account of prosodic phenomena. Current mainstream Autosegmental Metrical (AM) Phonology lacks such a functional component.

Furthermore, in view of multiparametric phonetic exponents in intricate function – form relations, discrete contrasts between prosodic units, with clearly marked boundaries between them, are the exception in speech communication rather than the norm postulated in AM phonology. This also casts doubt on the hypothesis of strict segmental anchoring [2] in the alignment of tonal targets with segmental boundaries, irrespective of rate or syllable structure. An f0-peak or an f0-valley contour-shift continuum across pre-accented, accented and post-accented syllables may, but need not, result in categorical perception in the Haskins sense of strong acuity across and weak acuity between category boundaries; yet the allocation to functional categories is always straightforward. There is category perception in all instances, but only some special cases are discrete categorical. The modelling of prosody needs to take this function – form relations into account.

2. The Kiel Intonation Model (KIM)

It has been shown for German in [3,4,5] that in f0-peak contour shifts the transition of the peak point into the accented vowel (from an early to a medial position) is perceived as a categorical change and allocated to a functional category change from Finality – ‘closing an argumentation’, to Openness – ‘opening an argumentation’. But the continuation of the shift towards the end of the accented vowel (to a late position) is perceived as a gradual change to the functional category of Unexpectedness – ‘contrast and expressive evaluation’. A comparable low-rising f0-valley contour shift produces no categorical perception, but there is nevertheless a category change from Casualness to Friendly Concern when the low valley point moves from an early position before the accented vowel onset to a late position in the accented vowel. The same function – form relations apply to English (cf. [6]), and no doubt to all West Germanic languages.

There are thus fundamental issues in the modelling of prosody within the traditional paradigm of structural linguistics. The ones that have been taken into account in the development of the Kiel Intonation Model are as follows.

• f0 patterns, as all other prosodic parameters, are related to categories of communicative function.

• Prosodic exponents identify communicative categories, without presupposing discrete boundaries for category differentiation.

• The short-time window of segment and the long-time window of prosody generation are synchronized in the coding of speech production and in the decoding of speech perception, with broad margins of segmental alignment around landmarks like accented vowels.

• f0-peak and valley patterns bear a holistic relationship to communicative categories and are not decomposed into local L and H tones of pitch accents, phrase accents and boundary tones at a formal prosodic phonology level.

• Although f0 is certainly the most powerful prosodic marker of communicative categories, other parameters, e.g. energy, long-term phonation, are also important contributors and need to be included in the analysis [7].
The analysis of f0-peak and valley synchronization has to include the shape of contours as a further factor in the identification of communicative functions [8]. They form a holistic bundle in category coding, such that, for example in German, an earlier f0-peak position combined with a slower fall from the peak maximum, after a faster rise to it, still signals a peak type for coding the OPENNESS category, and the opposite combination of synchronization and contour shape holds for a peak type coding finality.

The latter point illustrates the importance of pitch relations in the f0 movement relative to accented-vowel onset: in German (and the other West Germanic languages) finality is coded by high-low, openness by low-high. If f0 stays high longer after a faster rise to a (higher) peak maximum, this strengthens high pitch in the first part of the accented vowel, even if the peak maximum is synchronized earlier, and this higher pitch is linked to the speech function openness. If f0 falls faster after a slower rise to a (lower) peak maximum, low pitch is strengthened in the first part of the vowel even if the peak maximum is synchronized later, and this lower pitch is linked to the speech function finality. This is another aspect of the frequency code [9].

Synchronization, contour shape, and maximum height are thus not independent factors in the production and perception of communicative functions; it is their convergence in generating different pitch dynamics around the accented vowel that codes the different functions. The communicative goal allows speakers to use more than one of these pitch contours relative to articulatory landmarks, to adapt to the situational and contextual constrains in communication and to show individual preferences for synchronization or shape [10]. Only the functional approach, which enables the grouping of diverse measurement data to fixed categories in communicative behaviour that lie outside the analysed data, can give this insight.

In a preliminary experimental investigation, [10] compared synchronization and shape of f0-peak contours in the Neapolitan Italian declarative/interrogative function with the data found for finality/openness in German. In the interrogative function, the peak maximum is synchronized late in the accented syllable, whereas the declarative function has early placement. These synchronization differences are accompanied by shape differences, but there was a subset of speakers who made very little use of synchronization and coded the opposition by greater divergence in shape. So, we find the same variability in the generation of a pitch-dynamics contrast among speakers of the two languages, but in Neapolitan Italian the high-low vs. low-high pitch pattern codes declarative/interrogative, for which the frequency code was originally developed.

Unfortunately, the authors do not mention the anchoring of their data in communicative function, although it clearly determined their data collection. In faithful reverence to the formal pivot of AM phonology, they use the pitch accent categorizations H±L* vs. H* for German and L±H* vs. L*±H for Neapolitan Italian as their points of reference instead. But they would not have been in a position to group the data of “aligners” and “shapers” to the same formal categories if they had not had the fixed functional denominators. Furthermore, the description of questions vs. statements in Neapolitan Italian as depending exclusively on different pitch patterns leaves unresolved the manifestation of all the other types of the interrogative function, e.g. (surprise) repeat questions (cf. 3.4).

This shows that it is high time for linguistics to develop a comprehensive framework of communicative functions for prosody research.

3. Towards a Framework of Communicative Functions

A fair amount of empirical work has already been done [1, 11, 12, 13, 14], and Bühler [15] provides the theoretical foundation for its integration in a comprehensive framework.

3.1. Karl Bühler’s Organon Model

This model links the linguistic sign to the sender, the receiver, and the factual world of objects and factual relationships. From this three-faceted concept of the linguistic sign follow the related functions of expression, appeal, and representation, which are integral components in speech communication, determining behavioural interactions in varying proportion and weight. Speakers’ expressions are attitudes towards the Listener and the Factual World in communicative settings, or emphatic evaluations, or emotions; appeal to the Listener is carried by commands, requests and questions; factual representation features in propositions and information structure.

3.2. Information Selection and Weighting Function

In verbal interaction communicators select information points, which they weight in relation to each other in the communicative situation. The languages of the world use a great variety of formal means to achieve this, e.g. structural devices, such as deviation from default declarative word order, cleft sentence constructions, lexical intensifiers, but also sentence accentuation by segmental lengthening (of vowel nuclei and/or initial consonants), and by pitch and energy patterns in specific syllables to make them salient for information highlighting of words and syntactic elements in declarative, interrogative, and imperative functions.

Traditionally this information selection and weighting has been studied in propositional meaning. It led to the categories of ‘Focus’ and ‘Presupposition’ or ‘New’ and ‘Given’ in a framework of information structure. ‘Focus’ came to be studied in just one formal feature, namely f0 patterns of pitch accents, which was initiated by the analysis of English [16], but was also applied to German and Dutch. In these West Germanic languages, ‘Focus’ can be freely marked by pitch. This does, however, not apply to a language like French.

In a contextualization of the famous English example from [16], we may get the following dialogue interchange:

Speaker A: Peter invited a few of his friends to a party in his flat. Mary came with Manny: double ‘Focus’ with perceptually equal peak contour accents, in typical sentence declination, on Mary and Manny (Fig. 1).

Speaker B: No, Anna came with Manny: single ‘Focus’ with a peak contour accent on Anna, Manny is either completely deaccented if Speaker B only wants to set Anna in contrast to Mary (Fig. 2a), or it is partially deaccented with a low-rising valley contour if Speaker B picks up Speaker A’s background information as relevant for the statement (Fig. 2b).
Prosody-oriented studies of ‘Focus’ have been extended to a large number of languages, including Mandarin Chinese, in declarative and interrogative contexts [17,18]. They have all been concerned with propositional meaning in information structure. But information structure needs to be differentiated from argumentation structure, which is developed by communicators in ongoing discourse, parallel to, and different from, propositional information, introducing the speaker’s attitudes towards the listener and the factual world. In the examples of Fig. 2, for instance, a conveys ‘contrastive focus’ in the sense that speaker B’s information contrasts with that of Speaker A in a factual way. But the comparison of a and b shows that in the latter, attitudinal argumentative contrast is added by later synchronization of a higher peak (see 3.3).

3.3. Argumentation Function

The Speaker develops information points, selected for an utterance, into an argumentation structure with reference to 4 categories:

(1) **FINALITY**
In a concluding argument, the speaker sums up his/her own apperception of a communicative outcome as being final and no longer debatable.

(2) **OPENNESS**
In an opening argument, the speaker indicates that s/he has observed, and become aware of, something requiring further interaction.

(3) **UNEXPECTEDNESS – CONTRAST**
In an opening argument, the speaker indicates with an overlay of contrast that observation contradicts expectation.

(4) **UNEXPECTEDNESS – EXPRESSIVE EVALUATION**
In an opening argument with an overlay of contrast, the speaker adds expressive evaluation to an unexpected observation.

In English and the other West Germanic languages, these argumentation categories are coded by differently synchronized pitch peak patterns (with ToBI notations): (1) early (H+L*), (2) medial (H*), (3) late-medial (L+H*), (4) late (L*+H), coupled with different f0 peak and acoustic energy maxima. Fig. 3 illustrates (2), (3), (4) in the English sentence *He used to be slim.*, collected in the scenario of two people looking at old photos, one of them saying *Here is an old photo of Ken.*, and the other commenting on Ken’s former body size. For further details, and illustrations of the functional categories in the corresponding German utterances *Er war mal schlank.* in the same situational contexts, see [1].

3.4. Declarative and Interrogative Functions

The primary reference of the declarative function is the factual world constructed through the speaker’s argumentation. The main formal devices are declarative syntax, and elliptic phrases (e.g. in response to questions), as well as pitch and other prosodic patterning. In non-tone languages, default utterance pitch tends to end low or falling. Concern for the listener intervenes as the speaker qualifies the validity of a statement by phrases of the type *I think, I suppose, may be,* or by final low rising pitch in the West Germanic languages, or by voice quality (breathy voice).
On the other hand, the primary reference of the Interrogative function is the Listener to whom the speaker makes an Appeal to respond verbally, but there are several subcategorizations, again depending on varying Appeal – Expression – Representation constellations:

(1) **INFORMATION QUESTIONS** ask for specific information referring to subject, object, place, time, modality
   (1.1) focusing on **FACT ORIENTATION**: matter-of-fact appeal for information
       (1.1.1) **_EXPRESSION OF CONTRAST**
       (1.2) focusing on **LISTENER ORIENTATION**: request appeal for information, **ATTITUDE OF FRIENDLINESS**
           (1.2.1) **_EXPRESSION OF CONTRAST**
   
(2) **POLARITY QUESTIONS** ask for a decision as to the truth value of a proposition or argumentation along a positive-negative polarity scale
   (2.1) focusing on **LISTENER ORIENTATION**: decision left open for the listener to make
       (2.1.1) **_EXPRESSION OF CONTRAST**
       (2.2) introducing **SPEAKER ORIENTATION**: prejudging the listener’s decision towards one pole
       (2.2.1) **_EXPRESSION OF CONTRAST**
   
(3) **REPEAT QUESTIONS** ask for repetition of what has been said, in order to confirm an information point or the truth value of a statement
   (3.1) **ENQUIRING AN INFORMATION POINT** by question-word question or question-word ellipsis
       (3.1.1) **_EXPRESSION OF SURPRISE**
       (3.2) **ENQUIRING THE TRUTH VALUE** of a dialogue partner’s statement on a scale from (a) **AGREEING** to (b) **QUERYING**
           (3.2.1) **_EXPRESSION OF SURPRISE**

Whereas (1) is coded uniformly across the languages of the world by question words, e.g. English who, what, where, when, why, how, the coding of (2) relies on a great variety of formal means, including syntactic structure (initial verb position in the West Germanic languages), question particles (e.g. in Mandarin Chinese [18] and in literary Russian) or prosodic patterns on declarative syntax (e.g. in Neapolitan Italian [19]). But irrespective of the coding of the polarity function by syntactic or lexical means in a language, prosody codes subcategorizations. This prosody effect also applies to information questions. In repeat questions, prosodic patterns are of prime importance in combination with a variety of formal structures ranging from question-word to declarative syntax to elliptic phrases. The functional network of interrogativity will now be illustrated with reference to its formal coding in English.

### 3.4.1. **The interrogative function in English**

In all three basic types of interrogative functions prosodic patterns combine with formal syntactic structures to code the various categories of questions.

**Category (1) INFORMATION QUESTIONS**

Question-word structure is combined with nuclear peak patterns for **FACT ORIENTATION** (1.1), but with nuclear valley patterns for **LISTENER ORIENTATION** (1.2). The peak contour has medial synchronization with the accented vowel (see Section 3.3; ToBI: H*L-L%). For the **EXPRESSION OF CONTRAST** (1.1.1), it has late-medial synchronization (see Section 3.3; ToBI: L-H*L-H-L%). The valley contour is low-rising (or falling-rising) and has late synchronization with the accented vowel (ToBI: (H+J*L-H-H%). If the valley contour has early synchronization and is high-rising (ToBI: H*H-H-H%) the question-word structure no longer conveys an INFORMATION QUESTION but is a **REPEAT QUESTION** (3.1). The 4 different information question types are illustrated by the following dialogue.

**Speaker B:** "Where?"

**Speaker A:** “We’ll meet in Auchterarder tomorrow.”

(a small town in Scotland, not widely known).

(1.1) With a **medial peak** B asks for more information about the location of the venue in the town. This introduces the functional meaning of **OPENING ARGUMENT** (cf. 3.3) into the question context. (Fig. 4a)

(1.1.1) With a **late-medial peak** B stresses the need for more information about the venue against the insufficiency of the information so far given by A. This superimposes the functional meaning of **CONTRAST** on the **OPENING ARGUMENT** (cf. 3.3), and introduces it into the question context. The utterance has a tone of irritation and impatience: “But where? Your information is rather imprecise.” (Fig.4b)

(1.2) With a low (falling-rising **late valley**), where the rise starts in the accented vowel, B still asks for more information about the venue, but makes a request appeal to the listener. The fall adds contrast (1.2.1), but the utterance sounds less categorical and more friendly than with a peak pattern. (Fig.4c)

With a high-rising **early valley**, where the rise starts before the accented vowel, Speaker B appeals to the listener to repeat the name of the place because s/he has not heard properly or finds it strange (interrogative function (3.1), Fig.4d, cf. **Category(3) REPEAT QUESTIONS**).

![Figure 4: Speech waves, spectrograms, F0 (plain) and energy (dotted) of 4 interrogative functions in the question-word structure “Where?”](image)

**Figure 4:** Speech waves, spectrograms, F0 (plain) and energy (dotted) of 4 interrogative functions in the question-word structure “Where?”: a *medial peak* of type (1.1), b *late-medial peak* of type (1.1.1), c *late valley* (low falling-rising) of type (1.2), d *early high-rising valley* of type (3.1); female speaker, SBE..

**Category (2) POLARITY QUESTIONS**

Interrogative word-order structure is combined with nuclear valley patterns in **LISTENER ORIENTATION** (2.1), but with nuclear peak patterns in **SPEAKER ORIENTATION** (2.2).

The valley contour is high-rising and has early synchronization with the accented vowel (ToBI: H*H-H-H%).
For the **expression of surprise** (2.1.1), the high rise has late synchronization (ToBI: $L^\ast +HH-H$%)

The peak contour has medial synchronization with the accented vowel (see Section 3.3; ToBI: $H^\ast L-L$%). For the **expression of contrast** (2.2.1), it has late-medial synchronization (see Section 3.3; ToBI: $L+H^\ast L-L$%). These 4 different question types are illustrated by the following dialogue and Fig. 5; see also the exactly comparable data in the corresponding German utterances of [1].

(2.1) With a high-rising **early valley** where the rise starts before the accented vowel, A does not prejudge the answer but appeals to the listener for a polarity decision.

(2.1.1) With a high-rising **late valley**, where the rise starts in the accented vowel, A still appeals to the listener for a polarity decision, but with an expression of surprise at the person perhaps being in Rome.

(2.2) With a **medial peak**, A wants more information about the person’s whereabouts and suggests a place, expecting the answer to be “yes”. This introduces the function of **opening argument** of a medial peak contour into the question context.

(2.2.1) With a **late-medial peak**, A wants more information, as in (2.2), but contrasts his/her suggestion with his/her expectation. This introduces the superimposed functions of **contrast** and **opening argument** of a late-medial peak into the question context.

**Category (3) Repeat Questions**

In **inquiring an information point** they occur with question-word structure and a high-rising valley pattern. It starts on the question word and has early synchronization (3.1), see Fig. 6a (cf. also Fig. 4d); late synchronization adds the **expression of surprise** (3.1.1). The latter may be further heightened by breathy phonation, increased duration, increased f0 rise and increased energy for **negative intensification**, see Fig. 6b, cf. Section 3.5. For example,

Speaker B: “Where?” in the dialogue context of
Speaker A: “We’ll meet in Auchterarder tomorrow.”, cf. 3.4.1

(3.1) With a **early high-rising valley** of type (3.1),  
(3.1.1) With a **late high-rising valley** of type 3.1.1: expression of ‘surprise with negative intensification’, female speaker, SBE.

(3.2) In **inquiring the truth value** of a dialogue partner’s statement, a high-rising valley pattern is used in declarative syntax. It has early synchronization (3.2), Fig. 7a, or late synchronization for the additional **expression of surprise** (3.2.1), Fig. 7b. The latter may again be heightened in the same way for **negative intensification**, Fig. 7c, cf. 3.5:

Speaker B: “He is in Rome?” in the dialogue context of
Speaker A: “He has gone to Rome.”

(3.4.2) **Sequential restriction on questions in dialogue**

Since type (3.2) **repeat questions** take up a dialogue partner’s pronouncement as a whole, and enquire its truth value, they cannot ask for a polarity decision on new, additional aspects connected to it. So, “He is in Rome?” would not follow “He has gone to Italy.”, unless the speaker takes it for granted that the person would be in Rome when in Italy. The **polarity question** “Is he in Rome?”, types (2.1), (2.1.1), is the usual enquiry in this situational context.
The same function—form relations apply to English. Thus, “Where?” as an INFORMATION QUESTION with the EXPRESSION OF CONTRAST (1.1.1) may be given NEGATIVE INTENSIFICATION of type (a) with falling pitch. It expresses exasperation over not getting precise information. “Is he in Rome?” as a POLARITY QUESTION with an EXPRESSION OF SURPRISE (2.1.1) may also be given NEGATIVE INTENSIFICATION of type (a), this time with a late valley contour. The same applies to REPEAT QUESTIONS “He is in Rome?” and “Where?” (3.1.1), all expressing incredulity.

Fig. 8 compares “In a handbag.” as a simple Repeat Question (RQ), a Repeat Question with the Expression of Surprise and Negative Intensification (NI), “Lady Bracknell style”, and a Repeat Question with the Expression of Surprise and Positive Intensification (PI) (cf. [21]).

4. Applying the Framework in Comparative Prosodic Research

The framework of communicative functions outlined in 3, is a theoretical construct of human interaction, based on observation, which can be taken as a basic postulate for speech communication in any language. Linguistic analysis then needs to find out how the functions are formally manifested in the languages of the world. Thus, the functional framework becomes a powerful tool in comparative prosodic research.

4.1. Mandarin Chinese

Since tonal features are tied up in the lexical tones of a tone language like Mandarin Chinese, it is an empirical question how the categories of such a functional framework are implemented by its speakers and cognitively processed by its listeners in speech communication.

4.1.1. The Argumentation Function

With regard to the ARGUMENTATION function, the extensive data analysis of ‘Focus’ in declarative and interrogative structures in isolated sentences of Mandarin Chinese [17,18] is not sufficient to gain insight into communicative, as against linguistic, functions. It has only dealt with information structure in propositional meaning and cannot give answers about argumentation structure. Preliminary observations within the communicative perspective have been presented in [1] for
hao-3, xing-2 “OK” (recorded by Yi Xu, UCL London - Fig. 9, and Aoju CHEN, MPI Nijmegen - Fig. 10). As suggested by Aoju CHEN, they may be contextualized as follows:

**FINALITY**

Your boss asks you to hand in a project proposal soon. You explain that this is not feasible for various reasons. But your boss insists that your company needs the proposal quickly.

Your boss: Will you then hand in the proposal in two weeks?

You: hao-3, xing-2 (Figs. 9a,c, 10 a,c)

**OPENNESS**

Your boss asks you to hand in a project proposal soon. Because it is a very short proposal and you have a clear idea of what it should be like, you think it can be done quite easily.

Your boss: Will you hand in the proposal in two weeks?

You: hao-3, xing-2 (Figs. 9b,d, 10 b,d)

**Figure 9 upper panel (male speaker Yi XU)**

**Figure 10 lower panel (female speaker Aoju CHEN)**

Speech waves, spectrograms, F0 (plain) and energy (dotted) traces of Mandarin Chinese “hao-3” a F1 b OP and “xing-2” c F1 d OP.

For FINALITY (FI) the low tone ends low and concomitantly acoustic energy is low or decreases, whereas for OPENNESS (OP), pitch rises and energy is high or increases. Similarly, the high tone ends lower or higher and concomitantly the time course of acoustic energy is on a lower or higher level. In the FI context, English “OK” is realized as an early peak; f0 falls rapidly in the accented second vowel to a low level from high in the preceding unstressed vowel. In the OP context, it has a medial peak; f0 rises into the accented vowel from a medial level of the unstressed vowel, and then falls. FI strengthens low pitch, OP high pitch in the accented syllable (see 3.3). So, in Mandarin Chinese pitch lowering or raising is at work to code the argumentation functions of FINALITY and OPENNESS as it is in English (and German), simply adapted to the conditions set by the tone language.

4.1.2. **The Interrogative Function**

Questions have been studied extensively, but always from the angle of linguistic form. Even the detailed analysis of parallel encoding of interrogative meaning, focus and lexical tone in Mandarin Chinese by Liu and Xu [18] takes formal question types as the point of departure. Subjects read isolated sentences, graphically marked, among others, as statement by a period, ma particle question or so-called yes-no question, i.e. in declarative syntax with a question mark. This leaves out the essential situational embedding of different types of interrogative function in speech communication. Moreover, asking each speaker to read 380 sentences out of context cannot guarantee a reliable separation of the functional question types, although regularities of the influence on lexical tone patterns have been clearly demonstrated.

Fig. 11 provides instances of statements, ma and declarative-structure questions, which I collected from a female native speaker (Xiaojun ZHAO, 38yrs), reading the authors’ 4 tonal frames from Chinese script, marked by final period or question mark. In each tonal sequence, the register is raised for both question types across the whole sentence vis-à-vis the statement. In the rising sequence, the last rising tone additionally goes up very high, but this high rise is curtailed by the final neutral-tone ma. In the low-tone sequence, the last low tone falls and ends in creak in the statement, but rises high in the questions, also before the neutral-tone ma. In the falling sequence, the last falling tone ends also at a higher level in the questions, which is particularly striking when compared with the tone at the internal phrase boundary before a pause.

In all cases, the questions have a higher articulation rate than the statements. This is a phenomenon that was also found in German, see [20]. After an internal phrase break, there is tonal reset, at the statement or question register level, which is more marked in the questions.

So, it is not just the declarative-structure question that has higher pitch marking, spread across the whole utterance, to express interrogativity, but this applies to the ma particle question as well. However, these data do not give any information about the functional difference between the two question types, nor about the interpretation the speaker gave to their orthographic form when asked to read the communicatively strange sentences out of plausible context.

This is what we now need to find out. We want to know how the formal realization of the different INTERROGATIVE functions, superimposed on lexical tone in Chinese, compares with the function – form relationship found in the West Germanic languages. To this end, I collected some preliminary data from the same speaker by devising short English dialogues between a husband and wife who are wondering how to occupy their boy. The contextualization gave a high probability of eliciting POLARITY and REPEAT questions, with and without SURPRISE, as well as statements including CONTRAST.

The script was translated into Mandarin Chinese by the speaker and her French husband Alexsis Michaud, CNRS. Some of the dialogues are presented overhead; the target utterances are underlined.
(Tone 1 High)
Zhāng Wēi dānxīn XīāoYīng kāicē jīyuān
ZhangWei worry XiaoYing driving dizzy
‘Zhang Wei worries that XiaoYing will get dizzy while driving’

(Tone 2 Rising)
Wáng Méi huáiyí LiúNíng huáchuán zhāomi
WangMei suspect LiuNing canoeing obsessed
‘WangMei suspects that LiuNing will get obsessed with canoeing’

(Tone 3 Low)
LǐMín fāngān Lǐ-Yū ādiānhuǒ qīnuān
LiMin dislike LiuYu light a fire keep warm
‘LiMin dislikes LiuYu to light a fire to keep warm’

(Tone 4 Falling)
Yèliàng hǎipà ZhàoLǐ shūijiāo zuòmèng
YeLiang afraid ZhaoLi sleep dream
‘YeLiang is afraid that ZhaoLi will dream while sleeping’

Figure 11: Spectrograms and F0 traces, in each of the 4 frames of high, rising, low and falling tone sequences: in each panel, statement (top), declarative-structure question (centre), ‘ma’ particle question (bottom); a transcription window is linked to the latter (female speaker Xiaojun ZHAO).
Figure 12: Spectrograms, F0 (plain) and energy (dotted) + transcription for “Tā xiǎn qù hǎitān (ma)”: “ma” polarity question A (top), statement B (centre), prosody repeat question C1 (bottom) (female speaker Xiaojun ZHAO).

Figure 13: Spectrograms, F0 (plain) and energy (dotted) + transcription for “Tā (bu shì gèng) xiǎn qù hǎitān (ma)”: reinforced “ma” repeat question C2 (top), contrastive statement B1 (centre), prosody surprise repeat question D1 (bottom) (female speaker Xiaojun ZHAO).

Figure 14: Spectrograms, F0 (plain) and energy (dotted) + transcription for “tā xǐhuan qù dōngwùyuán (ma)”: “ma” polarity question A (top), statement B (centre), prosody repeat question C1 (bottom) (female speaker Xiaojun ZHAO).

Figure 15: Spectrograms, F0 (plain) and energy (dotted) + transcription for “tā (bu shì gèng) xǐhuan qù dōngwùyuán (ma)”: reinforced “ma” repeat question C2 (top), contrastive statement B1 (centre), prosody surprise repeat question D1 (bottom) (female speaker Xiaojun ZHAO).
Husband: What shall we do with the boy today? What do you think? A Does he want to go to the zoo/beach?

Wife: I don’t think so. B He wants to go to the beach/zoo.

Husband: C repeat questions requesting confirmation C1 He wants to go to the beach/zoo? C2 He does not prefer the zoo/beach?

Wife: D repeat question with disbeliefing surprise D1 He wants to go to the beach/zoo? I can hardly believe that.

The dialogues were read by the speaker in both roles of wife and husband. Future systematic data collection will have to record proper dialogues with pairs of a male and a female speaker and also swap the texts in the two roles so that complete data sets will be obtained from all female and male speakers. In the dialogue embedding, A is a polarity question, whether it is of the open, listener-oriented type (cf. (2.1) in Section 3.4) is not certain; B is a statement, which may also be contrastive B1 (cf. (2) and (3) in Section 3.3); C1 is a repeat question requesting confirmation (cf. (3.2) in Section 3.4); C2 reinforces the confirmation request of the preceding repeat question (cf. Section 3.5); D1 is a repeat question with disbeliefing surprise (cf. (3.2.1) in Section 3.4).

Figs. 12 and 13 compare non-emphatic and emphatic functions, respectively: ma questions A and C2 (polarity and reinforced repeat), statements B and B1 without and with contrast, prosody questions C1 and D1 (repeat and surprise repeat) in Tài bá shì gèng xǐhuan qù hàitān (ma). Figs. 14 and 15 do the same for Tài bá shì géng xǐhuan qù dōngwùyuán (ma).

Prosody repeat questions C1 differ from statements B by having their complete tone sequences shifted upwards in both sentences. Moreover, final high-tone âm rises instead of being level, and final rising-tone yuán rises continuously very high instead of evening out at a lower level. These data tie in with the statements and the prosody repeat questions of the 4 tonal frames in Fig. 11.

In the prosody surprise repeat question D1, the falling-rising f0 contour of low-high-tone hàitān is raised, and the f0 contour of falling-rising-tone dōngwùyuán is expanded in its low turning and its high end point, compared with the repeat question C1. In addition, the final syllable in both cases of D1 ends in breathiness.

When the statement is made by contrasting the communicatively focused word of B with the one in question A, the final syllable of hàitān or dōngwùyuán is given greater prominence than in the non-contrastive statement, by a 20-30% increase of its duration, by increased acoustic energy, and by tense phonation. Contrastive hàitān also has a full final nasal consonant with descending f0, resulting in a dome-shaped contour for the high syllable tone, which is level in the non-contrastive case.

The examples indicate that a raised register feature, superimposed on the lexical tone sequence of a declarative phrase structure, signals a repeat question. Expanding the f0 contour of the focused word and overlaying it with breathiness adds the meaning component of disbeliefing surprise. Increasing the prominence of the focused word adds contrastiveness to the statement.

The ma polarity question A differs from the statement B by both syllables of hàitān being raised, and f0 then descending slightly for the neutral-tone ma; in dōngwùyuán the low point of the fall and the high point of the rise are raised, f0 levelling out in ma. These patterns differ from the ma questions in the corresponding frames of Fig. 11, where all syllable tones are shifted upwards. This may be due to the latter being realized as listener-oriented open polarity questions in context-free isolated sentence productions, whereas the contextualization in the dialogues may have triggered the speaker-oriented reading “I suggest hàitān dōngwùyuán, what do you think?” This is a point for future research.

In the ma repeat question C2, the speaker reinforces his request for confirmation C1 by negating the object of his polarity question A. Compared with the polarity question, hàitān ma in the repeat question is shifted up under focus, and the slight dome shape contour is greatly expanded; dōngwùyuán has its fall and its rise expanded.

Finally, all the emphatic functions in Figs. 13 and 15 have a higher level of acoustic energy compared with the non-emphatic ones in Figs. 12 and 14.

4.1.3. Comparing Interrogative vs. Declarative Functions in Chinese with those in English

In Mandarin Chinese as well as in English raised pitch plays an important role in the signaling of repeat questions. Combined with declarative syntax, the pitch register of the entire question sentence is raised, vis-à-vis the statement, in Mandarin Chinese, while in English the main pitch effect resides in the final high rise starting on the last accent. When surprise is added to the request for confirmation, English uses a late, instead of an early, high-rising valley pattern, Mandarin Chinese expands the f0 contour of the tones in the focused word. In both languages, the surprise component may be further intensified, among others, by breathiness (see 4.1.4).

Polarity questions in English are word-order questions with either a nuclear fall or rise, depending on whether the speaker prejudices an affirmative answer or leaves the decision entirely to the listener. In Mandarin Chinese, polarity questions are ma particle questions. The collected data point to two different sentence intonation patterns: either the pitch register of the entire sentence is raised, compared with the corresponding declarative-syntax statement, or only the pitch contour of the tones in the focused word is raised. Whether this pitch extension difference reflects the same functional difference that is coded by pitch direction in English is a question to be solved by further data acquisition and analysis.

With C2, the speaker renews his request for confirmation immediately after his first repeat question C1, by negating the boy’s preference for the other locality. He may do this either by insisting on querying the truth value, or by requesting a response as to whether the negation is true. In the former case, the question form would simply be repeated in both languages. In the latter case, which was realized in the present data collection, Mandarin Chinese uses a ma particle question with reinforcement of the focused word. In English, the same function may be coded by a high register shift of an entire nuclear-fall pitch pattern (see 3.4.2), with the possible addition of the tag question does he? as a post-nuclear tail.
The contrastive statement B1 is realized by similar means in the two languages to increase the prominence of the focused word. Apart from accented-syllable lengthening, the f0 peak contour is expanded in English.

4.1.4. The Intensification Function

The prosody surprise repeat questions D1 and the reinforced ma repeat questions C2 exemplify negative intensification and reinforcement in questions, as illustrated for German and English in 3.5. In spite of the large-scale study of ‘Focus’ in Mandarin Chinese, e.g. [18], the systematic analysis of the Intensification function under focus has been largely excluded. Linguists across the world’s languages have shied away from investigating this function because belonging primarily to attitudinal and expressive meaning it has not been deemed to be a proper topic in formal linguistics and has therefore been relegated to paralinguistics as part of the study of emotion. However, this is the field where communicative functions are very likely coded, to a large extent, by the same prosodic means in the languages of the world. The use of breathiness in the Mandarin Chinese examples D1 of Figs. 12, 14 and in the corresponding intensified English repeat question of Figs. 6b, 7c, 8b bear witness to this.

4.2. The Frequency Code

Since higher pitch, either higher register or rising, serves to differentiate questions from statements in both Mandarin Chinese and English, adapted to the presence or absence of lexical tone constraints, it is legitimate to ask whether we are dealing here with a universal feature in all languages. Ohala [9] answered in the affirmative by proposing the Frequency Code, which means that in asking a question a speaker subordinates to a listener, and subordination is coded by a universal mechanism of raising pitch in speech. This version of the frequency code can certainly be regarded as an ingredient in coding and decoding questions. But when the subcategorization of questions is taken into account it becomes doubtful that Ohala’s Frequency Code can provide a general explanation for all types of questioning.

The function of repeat questions, without or with surprise, is to attract attention and stimulate a dialogue partner to action, rather than a passive subordination to the partner’s response. And polarity questions in which the speaker pre-judges the answer tend to have falling pitch in West Germanic languages. Furthermore, closer inspection of the interrogative function in speech communication has demonstrated the need for a subtle differentiation of many systematically determined types of questions which it is impossible to subsume under the same production principle.

It was, for instance, shown in [22] for German that the precursor to an accentual high rise may also be raised to a higher level. In a perception experiment using the Semantic Differential technique, listeners judged repeat questions with high precursors as expressing a more agreeable attitude towards the addressee (on a contrary – agreeable scale) than repeat questions with low precursors. The same may be expected for English. As the high rise in the repeat question has an activation function, this stimulation is intensified if the difference in pitch between the utterance beginning and the end of the rise is increased, i.e. by a low precursor, whereas the high precursor softens the repeat activation, making it more accommodating. In German and English, repeat questions may also have falling pitch patterns in declarative syntax transposed to a higher register to elicit confirmation that what is enquired IS true (see 3.4.2 and 4.1.3). This differs from the high-rising repeat question, which enquires the truth value one way or the other.

These subtle differentiations of communicative question functions need a more sophisticated Interrogative Code, which goes beyond the Frequency Code in trying to explain more than the linguistically triggered broad dichotomy of question and statement. The Interrogative Code encapsulates the multifarious relations in communicative interaction between Speaker and Listener where the listener-oriented Appeal function is central, but gets adjusted in various ways by the speaker’s attitudes and expressiveness. There seems to be overwhelming evidence that the Appeal function is coupled with high pitch [23,24], but the coexistence of both high rising and high register in the same language to code different question types makes it mandatory to develop a universally valid Interrogative Code cautiously within a framework of communicative functions.

Moreover, Rialland [25] has provided data from African register tone languages of the Sudanic region that do not show the high-pitch link with questions. Various combinations of an open vowel question marker, sonorant segment lengthening, delayed falling intonation and breathy termination occur to signal a question on the same syntactic structure as a statement. She refers to these property bundles as a ”lax prosody”. However, it is not clear what position these questions have in the framework of interrogative functions that has been presented in this paper. Most examples are translated as statements vs. repeat questions, e.g. beans, beans?, a slave, a slave? It is to be expected that these languages also have the functional category of polarity question and use different prosodic means to signal it.

5. Conclusion and Outlook

On the basis of detailed function-oriented data acquisition and prosodic analysis of argumentation structures, questions and emphasis in German and English it has been possible to propose a communicative framework of Argumentation, Declarative and Interrogative, and Intensification functions which control the formal exponents of speech interaction in these languages. The functions have been regarded as basic in human communication with reference to Bühler’s Organon Model, allowing their extrapolation to other languages as a general communicative frame into which each language fits its own formal means for coding the inter-language functions. This “function first, form second” approach, in turn, makes it possible to use the language-independent framework of communicative functions as a powerful tool in comparative prosodic research. However, its application necessitates a fundamental change in data acquisition methodology, because reading off isolated sentences will no longer do; tightly contextualized utterances in plausible texts are needed that can capture the functions to be investigated.

In a first step, the functional framework has been applied to the analysis of some formal exponents of the three functional domains in Mandarin Chinese, focusing on Declarative and Interrogative in greater detail. Data collection was based on situationally contextualized dialogues and contrasted with data from the reading of isolated sentences of systematically selected syntactic structures and lexical tone sequences. But the acquisition procedure still lacked the proper simulation of interaction between dialogue partners. Subsequent investigation will have to refine the
methodology of devising dialogue scripts and of implementing them in recording sessions with speaker pairs, and also needs to collect a lot more data from many more speakers. Moreover, the three functional domains have not received full coverage in Mandarin Chinese as yet. The argumentation categories finality and openness, with overlays of contrast and expressive evaluation, and the three types of intensification still require detailed study of their prosodic exponents. Even the interrogative function has not been examined exhaustively, although it has been dealt with more thoroughly than the other two domains.

Future analysis also needs to extend the proposed function—form investigation to a great variety of languages for insightful comparative prosodic research in speech communication. To provide an appropriate communicative basis for comparing varying formal devices, this cross-linguistic function—form analysis should, first of all, include those languages, e.g., Danish, French, Italian, Spanish, Swedish, where ‘Focus’ and ‘Sentence Mode’ have been studied quite extensively, mainly with isolated read sentence data (but cf. the functional approach to Swedish intonation in [26]). The comparative study needs to register tone languages of Africa as well, for which diverging question coding has been reported [25]. It will have to be ascertained what type of interrogative functions show this divergence and how the different types of interrogativity are manifested.

Such a cross-linguistic function—form approach will lead to a new, communicatively insightful prosodic typology of the world’s languages, and to a reassessment of traditionally postulated universal codes, like the Frequency Code, as generally valid aspects of prosody in the coding of specific speech functions. The theoretical framework of communicative functions will itself be further developed and refined.

This reinstates a goal Hermann [23, p. 390] formulated 70 years ago for future integral linguistic research of the world’s languages (“ganzheitliche Sprachforschung”): to capture “the distribution of different rhythmic-melodic properties over the different speech functions and their summation in one and the same function.” (“die Verteilung der verschiedenen rhythmisch-melodischen Mittel auf die verschiedenen Funktionen und ihre Summierung bei ein- und denselben Funktionen”).

6. Acknowledgements

I am very grateful to Aoju Chen and Yi Xu for making recordings for me, and to Fang Liu for stimulating exchanges on the analysis of questions in Mandarin Chinese and in English, following my reading of her PhD thesis. I am deeply indebted to Alexis Michaud and Xiaojun ZHAO for insightful discussion of my English dialogue scripts, for translating them into Mandarin Chinese and for recording them so competently. Any inadequacy in the interpretation of the data is my own.

7. References


