CHAPTER 29

....

PROSODY AND MEANING

....

JUDITH TONHAUSER

29.1 INTRODUCTION

....

THE prosody of an utterance, which includes the grouping of the words of the utterance, the relative prominences of the words and the tonal movements, makes tremendous and diverse contributions to the meaning of the utterance. Prosody can convey the emotional status of the speaker, such as anger, disgust, fear or sadness (e.g. Brosch et al., 2010). Prosody distinguishes between speech acts: for instance, the American English declarative sentence *It's raining* is a request for confirmation when uttered with a rise at the end, but an assertion when uttered with a fall at the end (e.g. Gunlogson, 2002). Prosody can also influence whether an interrogative sentence like *Do you want to do the laundry?* is understood as seeking information, as an invitation or as a request (the illocutionary act), and how annoyed, authoritative, and polite the speaker is taken to be (the perlocutionary effect) (Jeong & Potts 2016). Prosody can even influence the truth conditions of utterances. Whether, for example, the relative clause *who was on the balcony* in utterances of *Someone shot the servant of the actress who was on the balcony* is taken to modify *the actress* or *the servant* depends on the prosody of the utterance (e.g. Speer et al., 2011). Prosody also influences the truth conditions of utterances with focus-sensitive operators like *only* (e.g. Jacobs, 1983; Von Stechow, 1990; Rooth, 1992; Beaver & Clark, 2008). Thus, an utterance of *Kim only serves Sandy Johnnie Walker* in which *Johnnie Walker* is the prosodically most prominent expression is taken to mean that Kim serves Sandy nothing but Johnnie Walker, whereas an utterance in which *Sandy* is the prosodically most prominent expression is taken to mean that Kim serves nobody but Sandy Johnnie Walker.

This chapter examines the contributions of prosody to meaning by considering experimental research on a category of meaning that has been central to research on prosody and meaning, namely information-structural focus. The study of information structure is concerned with the question of how the (morphological, syntactic, semantic, prosodic, etc.) properties of an utterance reflect the organization of the discourse in which the utterance occurs (see, e.g., Lambrecht, 1994; Krifka, 2006; Féry & Krifka, 2008). In this chapter, 'focus' refers to an information-structural property of parts of utterances and is defined on the basis of questions: an expression is a focus of an utterance if the content of that expression answers the (possibly implicit) question that the utterance addresses

(see, e.g., Roberts, 2012a; Beaver & Clark, 2008). In (1), for instance, where B's utterance addresses A's interrogative utterance, the proper name *Mary* answers the question denoted by the interrogative utterance and hence is the focus of B's utterance. In (2), the focus of B's utterance is the indefinite noun phrase *a tornado*.[1] In this chapter, brackets subscripted with 'F' ([]$_F$) identify focused expressions. This notation should not be taken to convey any commitments about the syntactic representation of focus (but see, e.g., Rooth, 1992). To avoid confusion, the term 'focus' is not used in this chapter to refer to a phonetic or phonological property of focused expressions (e.g. 'prosodic focus').

(1) A: Who saw the tornado?
     B: [Mary]$_F$ saw the tornado.   (adapted from Roberts, 2012b: 1910)

(2) A: What did Mary see?
     B: Mary saw [a tornado]$_F$.

Question-answer pairs like (1) and (2) are the main diagnostic for information-structural focus, regardless of whether focus is marked prosodically, morphologically or syntactically in a given language.[2] As these examples illustrate, foci can be expressions whose denotations are given, like the proper name *Mary* in (1), or new, like the indefinite noun phrase *a tornado* in (2). Some works, such as Katz & Selkirk (2011), reserve the term 'focus' for contrastively focused expressions, to the exclusion of expressions providing new information, as discussed later.

This chapter proceeds as follows. Section 29.2 provides a brief introduction to prosody and how experimental investigations can meet challenges to exploring the contributions of prosody to meaning. Section 29.3 examines how expressions that are information-structural foci are prosodically realized, and section 29.4 how prosodically realized foci are perceived and interpreted. Both sections 29.3 and 29.4 discuss data from a variety of languages, the influence of context, how focus size is distinguished and the role of experimental methods in exploring the prosody of focus. The chapter concludes in section 29.5 with a brief discussion of the contributions of prosody to semantic/pragmatic phenomena related to focus, such as contrastive topic and presupposition. Additional linguistic phenomena influenced by focus are discussed by Kim (Chapter 25 in this volume).

---

[1] What is referred to as focus here is referred to as 'rheme' in other parts of the literature (e.g. Halliday, 1985; Roberts, 2012b). Other authors define 'focus' on the basis of sets of alternatives (e.g. Rooth, 1992), givenness (e.g. Schwarzschild, 1999) or a combination of givenness and providing an answer to the Question Under Discussion (QUD) (e.g. Büring, 2007). Different definitions of 'focus' are discussed in Kim's Chapter 25, in this volume.

[2] Some languages have been reported to not use prosody to mark information-structural focus, including Wolof (Rialland & Robert, 2001), Yukatek (Kügler & Skopeteas, 2006, 2007), Chichewa Chitumbuka, and Durban (Downing, 2008), Nłeʔkepmxcin (Koch, 2008), Northern Sotho (Zerbian, 2006), and Kʼiche' (Yasavul, 2013; Burdin et al., 2015). For discussion of prosodic, morphological, and syntactic marking of focus see, for example, Féry & Krifka, 2008, and Zimmermann & Onea, 2011. For the hypothesis that information-structural focus is marked prosodically in all languages, see Roberts, 1998.

29.2 PROSODY: A BRIEF PRIMER

This section is a brief introduction to prosody and to challenges to exploring prosody and meaning. For more detailed introductions to prosody from different perspectives see, for example, Pierrehumbert & Hirschberg, 1990; Hirst & Di Cristo, 1998: chapter 1; Gussenhoven, 2002; Grice & Baumann, 2008; Ladd, 2008; and Beckman & Venditti, 2011.

Any utterance has phonetic properties that can vary independently of the expressions uttered: for instance, in English, the sentence *Mary saw a tornado* can be uttered with prosodic prominence on *tornado* without changing the contribution the expression makes to the truth conditions of the sentence. Prosodic prominence on *tornado* can therefore be used to convey (what is referred to as) a post-lexical meaning. Cross-linguistically, the phonetic properties of utterances that can vary independently of the expressions uttered in order to convey post-lexical meanings include fundamental frequency, duration (of syllables or pauses) and intensity. Languages differ in whether a particular phonetic property is used to convey post-lexical meanings: for instance, a rise in fundamental frequency that is associated with the stressed syllable of an expression is a correlate of a post-lexical (high tone) pitch accent in English (e.g. *tornado* means 'tornado' regardless of whether it is realized with a high or a low tone pitch accent), but such a rise may be a correlate of a lexical pitch accent in Swedish (e.g. *anden* means 'the duck' or 'the spirit', depending on the pitch accent with which it is realized) or of tone in Mandarin Chinese (where, famously, *ma* means 'mother' or 'hemp' depending on whether it is realized with a high tone or a rising tone, respectively). Languages also differ in the phonetic properties that have become conventionalized bearers of post-lexical meaning and are thus part of the phonology of the language. The phonetic and phonological properties of a language that are used to convey post-lexical meaning constitute the prosody of the language.

The prosody of a language can provide cues about the organization of the utterance and its relation to the context in which it occurs: these cues include phrasing (how the expressions of the utterance are grouped), prosodic prominence (the relative prominence of the expressions in the utterance), and tonal movements (the tune of the utterance). Examples of such information cued by prosody include (see section 29.1) the attitude of the speaker, the speech act of the utterance, the syntactic and semantic relationships among the expressions of the utterance, and the information structure of the utterance. In English, for instance, whether the relative clause *who was on the balcony* modifies *the servant* or *the actress* can be indicated by marking either one of the two noun phrases as the right edge of a phrase, by a combination of lengthening and hyperarticulation of segments preceding the edge, pitch movements associated with the edge and a pause (for detailed discussion see, e.g., Wagner & Watson, 2010; and Speer et al., 2011). Thus, the study of prosody (or 'intonation', as it is also sometimes called) is not concerned with phonetic or phonological properties of an utterance that determine the lexical meanings of the expressions uttered, but with those properties that convey other types of meaning, under a broad use of the term 'meaning'. For an introduction to the autosegmental-metrical framework of intonational phonology assumed by many of the works reviewed in this chapter see, for example, Beckman, 1996; Shattuck-Hufnagel & Turk, 1996; and Ladd, 2008.

Given the contributions that prosody makes to natural language meaning, theories of meaning must capture how speakers use prosody to convey meaning and how listeners perceive and interpret the prosody of the utterances they hear. Studying the contributions that prosody makes to meaning is challenging for a number of reasons. First, since not all phonetic and phonological properties of an utterance contribute to post-lexical meaning, research on prosody must tease apart those that do from those that don't. This task is complicated by the fact that the prosody of any language is complex, with multiple components, some of which have continuous expression in the speech signal (see, e.g., Wagner & Watson, 2010, Watson, 2010, on prominence in production and comprehension). Second, since the vocal folds and speech patterns of the speakers of a language vary along a number of dimensions, the phonetic properties that underlie the prosody of the language do not have any absolute value for fundamental frequency, duration or intensity in the language, but vary by age, sex, etc. (e.g. men's speech is typically characterized by lower fundamental frequency values than women's). There are thus no invariant phonetic correlates of information-structural notions such as focus (see, e.g., Féry, 2007) and even the prosody of utterances that convey the same meaning is highly variable, both within and across speakers of the same language (e.g. Baumann et al., 2006; Ito & Speer, 2006; Speer et al., 2011). In short, the mapping between the prosody of an utterance in any given language and the meaning of the utterance is not straightforward. Furthermore, there is to date no agreement about the nature of the interface between prosody and meaning (for discussion see, e.g., Prieto, 2015). Open questions include, for example, whether the interface is direct, or mediated by syntax (for a probabilistic perspective, see, e.g., Calhoun, 2010), and whether individual tonal movements or combinations thereof are the basic meaning bearing units of prosody (see, e.g., Gussenhoven, 2002, for discussion and relevant references). And, finally, as noted earlier, there is cross-linguistic variation in the prosodic systems of languages, even closely related ones (see, e.g., Hirst & Di Cristo, 1998; Jun, 2005, 2014).

Experimental investigations can help meet these challenges to studying the contributions of prosody to meaning. Perception and comprehension experiments can, for instance, help tease apart those phonetic and phonological properties of utterances that influence post-lexical meaning from those that don't. Production experiments designed to explore the prosodic realization of utterances with particular meanings can serve to identify speaker-invariant prosodic cues to those meanings. Finally, cross-linguistic comparison is facilitated by running the same experiments with speakers of different languages. This chapter illustrates advances made by experimental investigations into the contributions of prosody to meaning, in particular the information structural notion of focus.

29.3 THE PRODUCTION OF PROSODICALLY REALIZED FOCUS

This section is concerned with the production of prosodically realized focus. Section 29.3.1 introduces prosodic cues to focus across languages, and argues that the prosody of focus in a given language may be independent of the typological characteristics of the prosodic system of the language. Section 29.3.2 shows that contextual information about focus

influences the prosodic realization of focus, which has consequences for the question of whether the grammars of languages distinguish different types of focus, as discussed in section 29.3.3. The idea that differences in focus size are prosodically distinguished is motivated in section 29.3.4. Section 29.3.5 concludes with a discussion of methods used in research that investigates how focus is prosodically realized.

29.3.1 Prosodic realization of information-structural focus across languages

In languages that mark information-structural focus prosodically, focused expressions are generally more prosodically prominent than expressions that are not focused. Languages differ in the phonetic and phonological means to implement prosodic prominence. These means include the presence and type of pitch accents and boundary tones, duration of segments and pauses, fundamental frequency range, and the alignment of pitch targets to the segmental string (e.g. Jun, 2005; Ladd, 2008; Watson, 2010; Féry, 2013). In English, for instance, focused expressions are made relatively more prosodically prominent than non-focused expressions in the same phrase by realizing the focused expression with a pitch accent, a longer duration, an expanded pitch range and greater intensity (e.g. Cooper et al., 1985; Eady & Cooper, 1986; Eady et al., 1986; Terken & Hirschberg, 1994; Xu & Xu, 2005; Breen et al., 2010). In Jun's (2005) typology of prosody, English is a head-prominence language since prosodic phrase-level prominence is marked on the phrase head, which is identified as the head of the phrase through pitch accenting. In Paraguayan Guaraní (Tupí-Guaraní), also a head-prominence language, focused expressions are distinguished from non-focused ones by the type of pitch accent realized, by the steepness of the slope of rising pitch accents and by duration (Clopper & Tonhauser, 2011, 2013; Burdin et al., 2015; Turnbull et al., 2015).

Korean, unlike English and Paraguayan Guaraní, is an edge-prominence language, which means that phrase-level prosodic prominence is marked on the left or right edge of the prosodic phrase through boundary tones (the left edge, in the case of Korean; see Jun, 2000; Kim, 2004). A Korean utterance is typically phrased into accentual phrases and (larger) intonation phrases. An accentual phrase, which is slightly larger than a word, is marked by a particular tonal pattern. A focused expression is made relatively more prosodically prominent than non-focused expressions by expanding its pitch range, lengthening its initial segment, and by reducing the pitch range on the following expressions and the duration of the expression before and after the focused expression. Optionally, a boundary tone may be inserted before the focused expression and the following words may be dephrased; that is, realized in the same phrase as the focused word (e.g. Jun, 1993; Jun & Lee, 1998; Jun & Kim, 2008).

Jun (2005: 441) proposed that the way in which information structure, including focus, is prosodically realized in a given language may depend to some extent on the prosodic structure of the language. This proposal was investigated on the basis of a study of the prosody of noun, adjective and noun phrase focus in American English, Paraguayan Guaraní, K'iche' (Mayan), and Moroccan Arabic in Burdin et al. (2015). In contrast to American English and Paraguayan Guaraní, which are head-prominence languages,

Moroccan Arabic and K'iche' were described in Burdin et al. (2015) as head/edge-prominence languages, which means that they exhibit both head-prominence and edge-prominence marking (see Jun, 2014). Burdin and her colleagues found that, contrary to Jun's proposal, 'the prosodic realization of focus is orthogonal to the head-, head/edge-, and edge-prominence distinction' (p. 262): for instance, focused expressions were lengthened in American English, Paraguayan Guaraní, and Moroccan Arabic, but not in K'iche'. (See Burdin et al., 2015: section 2.5, for discussion of other literature that also suggests this conclusion.) In sum, languages that mark information-structural focus prosodically rely on a range of phonetic and phonological cues, and which cues a language relies on may be independent of the typological characteristics of the language.

29.3.2 The prosodic realization of focus is context-dependent

Question-answer pairs are fundamental in research on information-structural focus: since interrogative utterances determine the foci of their answers, question-answer pairs allow researchers to explore how foci are (prosodically, morphologically, or syntactically) realized. It has long been observed that answers must be congruent with the interrogative utterances they address (e.g. Paul, 1880, 1919; Von Stechow, 1990; Rooth, 1992), as shown on the basis of the question-answer pairs in (3): J's utterance in (3a), with *Turkish* prosodically prominent (as indicated by the capital letters), is congruent with M's question in (3a), hence judged to be acceptable in response to that question, but not congruent with M's question in (3b), hence judged to be unacceptable in response to that question. The reverse pattern holds for J's utterance in (3b), in which *David* is prosodically prominent.

(3) Mandy, Craige, and Judith are having lunch at a place that serves Turkish, Lebanese, and Irish coffee.
  a. M: What kind of coffee does David like?
     J: David likes [TURkish]$_F$ coffee.
  b. M: Who likes Turkish coffee?
     J: [DAvid]$_F$ likes Turkish coffee.

(adapted from Tonhauser, 2016: 951)

It follows from question-answer congruence that the focus marking of an utterance constrains the question addressed by the utterance, regardless of whether the question is realized by an interrogative utterance, as in the examples in (3), or whether the question is implicit, as in many naturally occurring examples (see, e.g., Roberts, 2012a; Beaver et al., 2017). The fact that what is focused in an utterance can be conveyed both by context (e.g. the interrogative utterances in (3)) and by prosody suggests the information-theoretic hypothesis that the prosodic realization of focus is enhanced when the context does not identify which expression is focused. This hypothesis was investigated by Turnbull et al. (2015): in their production experiments, speakers of American English and Paraguayan Guaraní played an interactive game in which they instructed a confederate (of the same language) to place a series of tiles depicting different objects into numbered boxes on a

game board. The speakers' utterances were either made in a context in which it was predictable from the other game pieces (visible to the speaker and the confederate) whether the adjective, the noun, or the entire noun phrase was focused, or in a context in which the focus was unpredictable from the other game pieces. For instance, when the English utterance *Put the orange flower in box two* was made in a context in which the other game pieces were flowers of a variety of colours, adjective focus was predictable from the visual context; when the same utterance was made in a context in which the other game pieces consisted of a variety of objects, including flowers, in a variety of colours, adjective focus was not predictable from the visual context.

Turnbull and his colleagues found that the prosodic realization of focus was enhanced in both languages when the visual context did not predict which expression was focused. For instance, when adjective focus was not predictable from context, focused English adjectives were longer, realized with higher fo peaks and more likely to be realized with a rising pitch accent than when adjective focus was predictable from context. In Paraguayan Guaraní, whether focus was predictable from context did not affect pitch-accent type or duration, but the fo slope of accented adjectives was steeper (regardless of whether the adjective was focused) when focus was not predictable from context than when it was. Thus, the prosodic realization of focus is affected in both languages by the contextual predictability of focus, but the influence of contextual predictability on prosodic realization of focus appears to be language-specific.

Context-dependence in the prosodic realization of focus was also illustrated in Beaver et al.'s (2007) study of second occurrence focus, a repeated focus: in B's utterance in (4), *vegetables* associates with the focus-sensitive operator *only*, but *vegetables* is given. Second occurrence focus is marked with brackets subscripted with 'SOF'.

(4) A: Everyone already knew that Mary only eats [vegetables]$_F$.
     B: If even [Paul]$_F$ knew that Mary only eats [vegetables]$_{SOF}$, then he should have suggested a different restaurant.
                                                                    (Beaver et al., 2007: 247)

Beaver et al. (2007) found that second occurrence foci are not realized with a pitch accent, but distinguished from unfocused expressions by a longer duration and a higher energy. Thus, whether a focus is realized in American English with a pitch accent depends on the context in which the focus occurs.

In sum, the prosodic realization of focus in a given language depends on the contextual cues to focus and the context in which the focused expression is realized.

29.3.3 Prosodic realization of other information-structural properties of foci

Focused expressions denote answers to the (explicit or implicit) questions addressed by the utterances in which they occur, but they may differ along other information-structural dimensions. In the examples in (5), for instance, the noun phrase *a coffee* is the focus of A's utterances, but the denotation of the noun phrase differs along other information-structural dimensions: the denotation is new information in (5a), it is previously mentioned

and contrasted with the other member of a contextually-given set, namely tea, in (5b), and it is used to correct an interlocutor's assumption in (5c).

(5) a. Q: What would you like to drink?
        A: I'll have [a coffee]$_F$, please.
     b. Q: Would you like to have a coffee or a tea?
        A: I'll have [a coffee]$_F$, please.
     c. At the breakfast buffet, A may have coffee or tea, and has just been given a cup of tea.
        A: I don't want tea, I want [a coffee]$_F$.

Given that focus is not the only meaning conveyed prosodically (see section 29.1), it is unsurprising that foci that differ in their information-structural properties may differ in their prosodic realization. The three studies discussed in the following illustrate this for foci that differ along the lines of the examples in (5).

In Ito et al. (2004), American English-speaking participants gave instructions to a confederate about how to decorate a tree, thereby producing sequences of utterances consisting of a colour adjective and an object-denoting noun, such as . . . *blue ball* . . . *blue house*. The noun of an adjective-noun expression was defined as contrastive if it was immediately preceded by an adjective-noun expression with the same adjective but a different noun, like *house* in . . . *blue ball* . . . *blue house* (likewise for adjectives). They found that adjectives and nouns whose denotations were contrastive (whether new or given) were more likely to be realized with a L+H* pitch accent[3] than adjectives and nouns whose denotations were (non-contrastive and) new.

In Breen et al.'s (2010) experiments, speakers uttered answers and listeners had to identify the question they were answering: for example, an utterance of *DAmon fried an omelet this morning* answers *Who fried an omelet this morning?* but not *What did Damon fry this morning?* (recall that capital letters indicate prosodic prominence). An expression was defined as contrastive in the presence of an explicit contrast set: while *Damon* in the just-mentioned congruent question-answer pair is not contrastive, it is taken to be contrastive when *DAmon fried an omelet this morning* is uttered in response to *Did Harry fry an omelet this morning?*. In an experiment in which speakers were given feedback about whether listeners were successful in identifying the question they were answering, speakers' utterances distinguished between contrastive and non-contrastive subjects, verbs, and objects, with contrastive expressions realized with longer durations and higher intensity than (given or new) non-contrastive ones. Surprisingly, non-contrastively focused words were produced with higher fo than contrastively focused ones, a finding which may be in contrast to that of Ito et al. (2004), given that L+H* pitch accents are typically realized with a higher fo than H* pitch accents (for discussion and references, see, e.g., Ito & Speer, 2008).

Katz & Selkirk (2011) also found differences in the prosodic realization of contrastive and (non-contrastive) discourse-new expressions. Their production experiment compared the realization of expressions like *that Modigliani* and *MoMA* in utterances like *So he would (only) offer that Modigliani to MoMA* in which both expressions were discourse-new, or

---

[3] The L+H* pitch accent is a high tone pitch accent that is preceded by a low tone. For an introduction to the Tones and Break Indices (ToBI) annotation system, see Beckman & Ayers (1997).

only one was discourse-new and the other one was contrastive. For Katz & Selkirk (2011), an expression was contrastive if it associated with a focus-sensitive expression like *only*. They found that contrastively focused expressions did not differ from discourse-new ones in terms of pitch accenting or phonological phrasing, but that contrastively focused constituents had a longer duration, a greater relative intensity and a greater fo movement than discourse-new ones in the same position. (For relevant studies in other languages, see, e.g., Krahmer & Swerts, 2001, on Dutch; D'Imperio, 2001, on Neapolitan Italian; and Baumann et al., 2006, and Braun, 2006, on German.)

These three studies suggest that prosody does not merely distinguish focused and non-focused expressions (section 29.3.1), but also distinguishes whether a focused expression denotes new or contrastive information. It is important to note, however, that the term 'contrastive' was defined differently in the three studies reviewed here: its definition involved an expression in the immediately preceding utterance (Ito et al., 2004), a question that the utterance was correcting (Breen et al., 2010), and association with a focus-sensitive expression (Katz & Selkirk, 2011). The fact that the studies explored three different (though possibly related) notions of 'contrastive' focus may very well account for the differences observed in the three studies in how prosody distinguishes contrastively and non-contrastively focused expressions. These studies thus highlight the importance of unambiguously defining the meaning categories whose prosodic realizations are explored. (For discussion, see also Kim's Chapter 25 on focus, in this volume.)

The findings of the three studies also bear on the question of whether foci that differ in other information-structural properties are grammatically distinguished. Katz & Selkirk (2011), for instance, argued that contrastive foci and discourse-new foci are represented differently in the syntax and thereby receive distinct interpretations. (For related proposals see, e.g., Kiss, 1998; Vallduví & Vilkuna, 1998; Selkirk, 2007.) An alternative position (which is compatible with foci being syntactically represented or not) is one on which differences between foci are attributable to context and hence need not be grammatically represented (see, e.g., Rooth, 1992; Roberts, 1998; Büring, 2007).

29.3.4 Prosodic marking of focus size

This section shows that prosody provides cues to the size of the focus of a given utterance. B's utterances of *Kim baked a cake* in (6a–c), for instance, differ in whether the focus is the direct object, the verb phrase, or the entire sentence, respectively.

(6)   a. A: What did Kim bake today?
          B: Kim baked [a cake]$_F$.
      b. A: What did Kim do today?
          B: Kim [baked a cake]$_F$.
      c. A: What happened?
          B: [Kim baked a cake]$_F$.

Since B's utterances in (6a–c) can all be realized with pitch accents on *cake*, and no other pitch accents, some authors assume that prosodic marking of focus is systematically

ambiguous with respect to whether the direct object, the verb phrase, or the entire clause is focused (e.g. Gussenhoven, 1983; Selkirk, 1996). Accordingly, Selkirk's (1996) widely adopted Focus Projection Hypothesis predicts that a pitch accented direct object can 'project' focus from the direct object to the verb phrase and to the entire utterance. (See Ladd, 2008: chapter 6, for a discussion of focus projection.)

However, experimental research on the prosodic realization of focus has provided empirical evidence that the size of focus can be prosodically distinguished. For American English, Breen et al. (2010) found that speakers produce subject-verb-object sentences differently depending on whether the entire sentence or only the object is focused. Specifically, they found that when only the object is focused, the object is produced with the highest maximum fo, the longest duration and the maximum intensity compared to other words in the sentence, whereas the words of the sentence were more prosodically similar to one another when the entire sentence was focused. Baumann et al. (2006) investigated the prosodic realization of German sentences with sentence focus, verb phrase focus, and object focus, and found that the narrower the focus was, the more likely the direct object was to be produced with a longer duration and with a higher fo peak relative to previous fo peaks, and the less likely the verb was to be realized with an accent. For European Portuguese, Frota (2002) found that the object is realized with a longer duration and a different pitch accent when only the object is focused than when the entire sentence is focused. Similarly, Hanssen et al. (2008) showed that, in Dutch, sentence and object focus are distinguished by the duration of the object and by the timing, scaling, and slope of fo targets and movements.

Focus size is also distinguished in languages genetically unrelated to European ones. For Korean, for instance, a verb-final language, Kim et al. (2006) found that verb phrase focus was distinguished from sentence focus by inserting an intonational phrase boundary, by raising the pitch peak of the first word of the verb phrase, and by lengthening the first and second words in the verb phrase. Jun & Kim (2008) observed that the object was relatively more prominent in object focus than in verb phrase focus: in object focus, the fo peak of the object was higher and the object was longer than in verb phrase focus, and the verb phrase was more likely to be dephrased or produced with a reduced pitch range in object than in verb phrase focus.

Finally, focus size is also marked within the noun phrase. As mentioned earlier, Burdin et al. (2015) compared the prosodic realization of focus in utterances in which the adjective, the noun or the entire noun phrase was focused across four genetically unrelated languages. For American English, they found that noun focus is distinguished from noun phrase focus in that the adjective is less likely to be unaccented when the noun phrase is focused than when the noun is focused. (For a similar finding in Dutch, see Krahmer & Swerts, 2001.) For Paraguayan Guaraní, in which the noun precedes the adjective, they found that adjectives were longer when the noun phrase was focused than when the adjective was focused. In Moroccan Arabic, the noun again precedes the adjective. Here, Burdin and her colleagues found that both the adjective and the noun were longer when the noun phrase was focused than when the adjective was focused. (K'iche' did not show any prosodic marking of focus; see footnote 2.)

In sum, genetically and typologically unrelated languages use a variety of prosodic means to lend prominence to particular parts of an utterance and to thereby identify the size of the focus of the utterance. These findings suggest that the assumption that underlies Selkirk's

Focus Projection Hypothesis, that prosodic focus marking does not distinguish focus size, may not be empirically adequate, at least from the production perspective. Section 29.4.4 addresses whether listeners attend to prosodic cues to focus size.

29.3.5 Experimental tasks to explore the prosodic realization of focus

Experimental research on the prosodic realization of focus relies on a variety of tasks: recordings are made of individual participants reading scripts (e.g. Beaver et al., 2007; Katz & Selkirk, 2011), of pairs of participants reading scripts to one another (e.g. Clopper & Tonhauser, 2013), of participants completing a task with a confederate (e.g. Ito & Speer, 2008; Burdin et al., 2015) and of pairs of participants engaging in an interactive task (e.g. Watson et al., 2008; Breen et al., 2010). Compared to empirical claims about the prosodic realization of focus that are based on individual researchers' intuitions, experimental investigations have the advantage of allowing researchers to quantify phonetic and phonological cues to focus, and to distinguish between speaker-invariant and speaker-specific prosodic cues to focus.

The aforementioned tasks differ significantly in the extent to which the elicited utterances resemble conversational speech, with participants reading scripts being the least and participants engaging in interactive tasks being the most like conversational speech. An advantage of scripted language is that researchers have full control over the content, length, and structure of the produced utterances, thus facilitating the comparison of productions of pairs of utterances that differ only in the variables of interest. However, as discussed in Ito & Speer (2006), one of the worries of investigating prosody on the basis of productions of read, scripted language is that since such productions differ from conversational speech, the results of the investigation have limited generalizability. Ito & Speer therefore urge researchers to rely on tasks that examine 'the relationship between produced intonation and the linguistic information conveyed in a discourse . . . without changing the nature of the speech people typically produce in conversation' (p. 233). See the references earlier in this section for examples of such tasks.

29.4 THE PERCEPTION AND INTERPRETATION OF PROSODICALLY REALIZED FOCUS

This section examines how listeners perceive and interpret prosodically realized focus.[4] Section 29.4.1 examines the perception and interpretation of prosodically realized focus

---

[4] The research reviewed in this section is limited to experimental research, to the exclusion of works that analyse the meaning of (mostly English) prosody by abstracting over findings in the experimental literature (e.g. Pierrehumbert & Hirschberg, 1990; Steedman, 2000, 2004, 2014; and Truckenbrodt, 2012). Specifically, the research reviewed in this section is limited to works based on perception tasks, which ask participants to make a judgement about the signal (e.g. prominence ratings), and interpretation tasks,

---

across languages, and 29.4.2 shows that the perception and interpretation of prosodically realized focus is context-dependent. Section 29.4.3 illustrates that information-structural differences between foci are perceived and interpreted, and section 29.4.4 addresses the perception and interpretation of focus size. Section 29.4.5 concludes with a discussion of methods.

29.4.1 The perception and interpretation of focus across languages

Not only do speakers use prosody to signal what is focused in an utterance (section 29.3.1), listeners also attend to prosodic cues to focus in their interpretation of utterances. In an early study on American English, Most & Saltz (1979) found that listeners rely on the location of pitch accents in utterances of sentences like *The pitcher threw the ball* in identifying whether the subject or the object is focused. In one of their experiments, participants listened to utterances of such sentences in which either the subject or the object was prosodically prominent and were asked to write a question to which the utterance would be an appropriate answer. Most & Saltz found that the expression that was prominent in the answer utterance corresponded, in 68 per cent of cases, to the expression that was focused in the utterance, given the question that the participant had provided. For instance, the question *Who threw the ball?* was written for the utterance *The pitcher threw the ball* in which the subject was prominent. Birch & Clifton (1995) also relied on question-answer congruence to explore whether listeners attend to prosody in identifying foci. Asking native speakers of American English to judge the appropriateness of congruent and incongruent question-answer pairs, they found that listeners rated utterances in which a focused noun phrase was pitch-accented as more appropriate than utterances in which a focused noun phrase was not pitch-accented. Breen et al. (2010) also found that listeners in their Experiment 2 (where speakers and listeners received feedback about whether the listeners' response was correct) were able to determine whether the subject, verb, or object was focused. In sum, these studies provide empirical evidence that listeners attend to prosodic cues in identifying which expression of an utterance is focused.[5]

Clopper & Tonhauser (2013) investigated whether Paraguayan Guaraní listeners attended to prosodic cues to focus, as well as which prosodic cues they attended to. As mentioned in section 29.3.1, Clopper & Tonhauser identified, on the basis of production experiments, that focused expressions in Guaraní are more likely than non-focused expressions to be realized with a rising pitch accent, with a longer duration and a steeper slope for rising pitch accents. The question-answer pairs collected in one of the production experiments served as stimuli for an interpretation and a perception task. In the interpretation task, Guaraní listeners were asked to identify which of two lexically identical answers

---

which ask participants to make a judgement about meaning (e.g. whether an utterance is congruent with a question).

[5] Accent placement facilitates listeners' utterance comprehension (e.g. Bock & Mazzella, 1983; Terken & Nooteboom, 1987; Ayers, 1996) and information about focus from accent placement is processed rapidly online (e.g. Dahan et al., 2002). For a more thorough discussion of the influence of prosodic prominence on processing and memory, see Kim's Chapter 25 in this volume.

that differed in their prosody was the preferred response to a question. They found that listeners attended to the pitch-accent type and the slope of rising pitch accents in identifying which expression in the utterance was focused. However, they also found that Guaraní listeners' performance was significantly lower than chance on this task: that is, listeners were not able to reliably use prosodic cues to identify which expression in an utterance was focused, contrary to the findings of Breen et al. (2010). As discussed in Clopper & Tonhauser (2013: section 4.6), this result may at least partially be attributed to the task design (listeners did not receive feedback about whether their answers were correct) and to the fact that the materials included highly variable natural Guaraní utterances that included incongruous prosodic cues to focus.

In the prominence perception task, Clopper & Tonhauser asked Guaraní listeners to identify which word of the two-word answer utterances was prosodically more prominent. They found that listeners attended to the duration of the stressed syllable in identifying which expression was perceived as prosodically more prominent. As in the interpretation task, performance was not significantly above chance: listeners judged the focused expression to be more prosodically prominent than the non-focused one in 52 per cent of cases. As discussed in Clopper & Tonhauser (2013: section 5.5), the chance performance is consistent with research examining the perception of prosodic prominence by naïve listeners in other languages (e.g. Streefkerk et al., 1997; Mo et al., 2008). Taken together, the results from interpretation and perception tasks suggest that prosodic cues to focus constrain but do not determine what is focused in an utterance.

In sum, there is evidence from genetically unrelated languages that listeners attend to prosodic cues in identifying what is focused in an utterance. Perception and interpretation experiments can also serve to identify which prosodic cues to focus listeners attend to, as well as how successful listeners are in determining which expression is focused from prosodic ones. Whether languages differ in the strength of their prosodic cues to focus is an open question.

29.4.2 Contextual influence on the perception and interpretation of prosodically realized focus

Context not only has an effect on the production of prosodically realized focus (section 29.3.2), but also on its perception and interpretation. Bishop (2012), for instance, showed that American English listeners' perception of prosodically realized focus is influenced by the question that the utterance addresses. In his study, participants listened to question-answer pairs, like the three shown in (7), and were asked to rate how 'stressed' each word in the answer utterances sounded. Crucially, the same recording was used for the answer in each question-answer pair, namely a recording of the answer with verb phrase focus.

(7) A: What happened yesterday? / What did you do yesterday? / What did you buy yesterday?
     B: I bought a motorcycle.

Bishop found that the objects of the answer utterances were rated as more stressed when the answers were heard in response to an object focus question than in response to a verb

phrase or sentence focus question. Thus, 'listeners' judgements of prosodic prominence are significantly and independently affected by their interpretation of an utterance's information structure' (p. 255): that is to say, by what is contextually identified as the focus of the utterance. Similarly, Turnbull et al. (2014) found that American English expressions that were realized with L+H* pitch accents were more likely to be perceived as prosodically prominent when listeners were explicitly made aware of the contrastive context in which the utterances were made.

The perception of prosodic prominence is also influenced by the prosodic context. In Krahmer & Swerts's (2001) study, Dutch listeners were asked to distinguish between words realized with (what the authors refer to as) contrastive and new accents. When presented with expressions realized with such accents in their sentential context, expressions with contrastive accents were rated as more prominent than expressions with new accents, but this difference in prominence rating was not observed when the expressions were presented in isolation. Thus, listeners attended to prosodic cues to the information-structural status of the expressions, but these cues were contextually enhanced.

Context also influences the interpretation of prosodically realized focus, as shown by Kurumada et al. (2012). Their study investigated the interpretation of prosodically realized focus using sentences of the form *It looks like a . . .*, such as *It looks like a zebra*. Utterances of these sentences differ in their interpretation depending on whether the verb *looks* or the noun (e.g. *zebra*) is the focus: while the utterances typically favour an affirmative interpretation ('it's a zebra') regardless of which expression is focused, the negative interpretation ('it's not a zebra') is more likely to arise when the verb is focused than when the noun is focused. In one experiment, Kurumada et al. (2012) showed that when listeners were made aware that speakers can also produce *It's a zebra*, they were more likely to assign negative interpretations to *It looks like a zebra*, regardless of focus placement. This finding suggests that contextual information about the speaker's intended meaning can weaken the contribution of prosodically realized focus to meaning. Another experiment reported in Kurumada et al. (2012) suggests that information about a speaker's use of prosody to convey meaning influences the extent to which listeners rely on prosodically realized focus in deriving the speaker's intended meaning.

In conclusion, the studies reviewed earlier show that the perception and interpretation of focus is context-dependent—at least in English and other well-studied European languages.

29.4.3 Interpreting prosodic cues to foci that differ in other information-structural properties

The studies reviewed in this section show that listeners attend not just to prosodic cues to focus (section 29.4.1) but also to prosodic cues to other information-structural properties of foci. Breen et al. (2010), for instance, found that a few American English listeners were able to rely on prosodic cues to distinguish whether an utterance like *Damon fried an omelet this morning* was made in response to a question like *Who fried an omelet this morning?* (in which *Damon* is new information) or to a question like *Did Harry fry an omelet this morning?* (in which *Damon* is corrective and contrasted with *Harry*). They found that more listeners were able to identify which question the utterance was responding to when the

utterance was preceded by *I heard that . . .* , which suggests that 'speakers are able to convey contrastiveness using words outside of the clause containing the contrastively focused element' (p. 1089).

Many studies have compared the interpretation of high tone pitch accents (H* in the ToBI annotation system, see footnote 3) and L+H* accents. In one of Ito & Speer's (2008) eye-tracking experiments, listeners heard sequences like . . . *green drum . . . blue drum* and were found to fixate on the relevant objects (e.g. drums) earlier when *blue* was realized with a L+H* accent than with a H* accent. This finding suggests that listeners are more likely to interpret adjectives as contrastively focused when they are realized with a L+H* accent than when they are realized with a H* pitch accent. Likewise, Gotzner & Spalek (2014) showed, using a truth-value judgement task, that German expressions were more likely to generate exhaustive inferences when realized with L+H* rather than H* accents. The question of whether H* marks foci whose denotation is new (as claimed in, e.g., Pierrehumbert & Hirschberg, 1990), was addressed in Watson et al.'s (2008) eye-tracking study. While they did find that American English listeners were more likely to fixate on a contrastive referent than on a new one when they heard the L+H* accent, listeners were not more likely to fixate on a new referent than a contrastive one when they heard the H* accent (an expression was taken to be 'contrastive' if it contrasted with the denotation of an expression in an immediately prior utterance, similarly to Ito & Speer, 2008). Watson et al. (2008) took these findings to show that L+H* conveys a contrastive meaning (in support of Pierrehumbert & Hirschberg, 1990) but that H* is compatible with new and contrastive foci.

Although the aforementioned studies suggest that listeners interpret H* and L+H* differently, other research suggests that the two pitch accents do not correspond to two distinct perceptual categories. Bartels & Kingston (1994), for instance, synthesized a continuum of stimuli between H* and L+H* by varying the acoustic characteristics of the accent (peak height, peak alignment, peak timing, depth of the dip preceding the peak) and the context in which the sentences were presented. They asked American English listeners to rate whether the relevant expression received a contrastive or a non-contrastive interpretation. (Here, contrastiveness was defined as the explicit denial of a salient alternative, like *apple* in *Amanda had a banana.—No, she had an apple.*) Bartels & Kingston did not find clear evidence for a categorical boundary between L+H* and H*. What they did find was that, the higher the peak height was, the more likely a contrastive interpretation was. (But see Calhoun, 2004, for a different finding.) The depth of the dip preceding the peak and the timing of the peak were secondary cues to a contrastive interpretation. Ladd & Morton (1997) found that listeners categorically distinguished the two accents in an identification task but they did not identify a category boundary in discrimination tasks, which they took to mean that the two accents are categorically interpreted, but continuously perceived.

29.4.4 Perceiving and interpreting the size of prosodically realized focus

Listeners can perceive and interpret prosodic cues to differences in focus size. D'Imperio (1997), for instance, found, using a prominence perception task, that Neapolitan Italian

listeners perceive utterances of subject-verb-object declarative sentences with sentence focus differently from utterances with object focus: listeners identified the object as most prominent in only 56 per cent of sentence focus utterances but in 77 per cent of object focus utterances.

Can listeners interpret prosodic cues to differences in focus size? Gussenhoven (1983) answered this question negatively. He found that participants were at chance in distinguishing whether an utterance of a subject-verb-object sentence, like *We repair radios*, that had been made in response to a question that elicited object focus (*What is it you repair?*) responded to that question or to a question about the verb phrase (*What is the nature of your business?*). Breen et al. (2010), on the other hand, found that American English listeners can distinguish between object and sentence focus: listeners in their Experiment 2 were able to correctly identify utterances that were produced with (non-contrastive) object focus 57 per cent of the time, perceiving such utterances as having sentence focus only 13 per cent of the time. Similarly to Gussenhoven's (1983) task, listeners in their study had to identify which question an utterance responded to, but in contrast to Gussenhoven's task, listeners in their study were seated in the same room as the speakers, and both the speakers and the listeners received feedback when a listener chose the wrong question: 'This cue was included to ensure that speakers knew when their productions did not contain enough information for the listener to choose the correct answer' (Breen et al., 2010: 1067). Since the speakers' utterances more clearly disambiguated focus size when such feedback was given, it is possible that the listeners' ability to perceive distinctions in focus size depends on whether speakers are aware of the need to disambiguate their productions.

As noted in Bishop (2017), which provides an overview of research into the perception and interpretation of prosodic cues to focus size, previous studies seem to suggest that prenuclear prominence (i.e. a pitch accent on the verb) may be 'less relevant to identifying broad focus than to identifying narrow focus' (p. 238). The results of his investigation into the processing of prosodic cues to focus size, which involved a cross-modal priming task, suggest that prenuclear accents were unexpected when the object of subject-verb-object sentences was narrowly focused, but did not influence reaction times when the verb phrase of the sentence was focused. While Bishop's findings are compatible with Selkirk's (1996) Focus Projection Hypothesis, studies like D'Imperio (1997) or Breen et al. (2010) challenge the assumption that an utterance with a prosodically prominent object noun phrase is ambiguous between object, verb phrase, and sentence focus.

29.4.5 Experimental tasks for the perception and interpretation of prosodically realized focus

Although researchers have intuitions about how they perceive and interpret prosodically realized focus, the preceding sections have shown that experimental investigations are essential to identifying the phonetic and phonological properties of utterances that naïve listeners attend to. Experimental investigations can also identify inter-speaker variation in the prosodic cues attended to; for discussion see, for example, Breen et al. (2010) and Bishop (2017). The preceding sections have also highlighted that research at the interface of prosody and meaning must carefully define the meanings that are investigated, including

'focus' or 'contrastive focus'. Finally, research must consider that the perception and interpretation of prosodically realized meaning, such as focus, may be context-dependent and control for context: when context is not controlled for, participants' interpretations may differ simply because participants imagine different contexts.

Experimental research on the perception and interpretation of prosodically realized focus employs a wide range of tasks: participants are asked to identify important or prosodically prominent words (e.g. D'Imperio, 1997; Clopper & Tonhauser, 2013), to identify congruent question-answer pairs (e.g. Most & Saltz, 1979; Gussenhoven, 1983; Birch & Clifton, 1995; Breen et al., 2010; Clopper & Tonhauser, 2013), or to answer questions about what a particular utterance means (e.g. Kurumada et al., 2012), or whether the utterance is true (e.g. Terken & Nooteboom, 1987). Eye-tracking and priming experiments provide information about the processing of prosodically realized focus (e.g. Dahan et al., 2002; Ito & Speer, 2006; Bishop, 2017). Stimuli can be resynthesized utterances (e.g. Ladd & Morton, 1997; Calhoun, 2004) or utterances produced by trained speakers (e.g. Gussenhoven, 1983) or untrained speakers (e.g. Breen et al., 2010; Clopper & Tonhauser, 2013); they may also differ in whether they realize specific prosodies (e.g. D'Imperio, 1997; Ito & Speer, 2006) or are randomly selected from a production experiment (e.g. Breen et al., 2010; Clopper & Tonhauser, 2013). See Watson et al. (2006) for arguments in favour of experimental tasks that can be used to explore both the production and comprehension of prosody, and tasks that allow for the integration of contextual information.

29.5 CONCLUDING REMARKS

In this chapter, an information-structural focus was defined as an expression that answers the question addressed by the utterance in which the expression is realized. As discussed in the preceding sections, the prosodic realization of utterances in many languages constrain what is focused in the utterance, and speakers of these languages can perceive and interpret prosodically realized foci. Experimental investigations are invaluable in identifying the phonetic and phonological cues to focus because such cues are subject to inter-speaker and inter-utterance variation, and because the phonetic and phonological properties of utterances are not only implicated in conveying focus. Although many open questions remain, the results of this research already have important implications for theories of meaning. For instance, theories of meaning must consider that the prosodic realization of focus is highly variable across utterances, speakers, and contexts. Similarly, although the prosodic realization of an utterance provides cues to focus, prosodically realized focus does not, typically, unambiguously determine the perception and interpretation of focus in such utterances, even when the prosodic cues to focus are consistent and speakers were aware of the ambiguity to be resolved. As a consequence, theories of meaning that start from the premise that the input to interpretation unambiguously marks what is focused in an utterance (e.g. via F-marking in the syntax) constitute an idealization of the complex interplay of prosody and focus. (Discrepancies between empirical findings and theoretical analyses in research on clefts are discussed in Onea's Chapter 24 in this volume.)

Information-structural focus is related to or implicated in other categories of meaning. As a consequence, studies of the production, perception, and interpretation of focus are relevant for a broad set of meanings beyond focus. Contrastive topics, for instance, have been defined as expressions that answer super-questions to the question addressed by the utterance in which the expression is realized; that is, as foci relative to a super-question (e.g. Jackendoff, 1972; Roberts, 2012a; Büring, 2003). It is therefore not surprising that contrastive topics, too, can be prosodically realized, and that prosodically realized contrastive topics can be perceived and interpreted (e.g. Braun, 2006; Wagner et al., 2013). Contrastive topic-marking in turn influences the interpretation of sentences with scalar adjectives: De Marneffe & Tonhauser (to appear) found that American English utterances with scalar adjectives that were realized with the contrastive topic contour were more likely to give rise to a scalar implicature than when the utterance was realized with (what they referred to as) a 'neutral' contour. More generally, research on conversational implicatures has established, for a range of scalar expressions, that the relevant scalar implicature is more likely to arise when the scalar expression is prosodically prominent than when it is not (e.g. Chevallier et al., 2008; Thorward, 2009; Zondervan, 2010; Cummins & Rohde, 2015; Schwarz et al., to appear). Lastly, it has been hypothesized that information-structural focus is implicated in presupposition projection (e.g. Beaver, 2010; Beaver et al., 2017; Simons et al., 2017), and experimental research on presuppositions has shown that the prosodic realizations of utterances with presupposition triggers influences the projectivity of the presuppositions (Cummins & Rohde, 2015; Tonhauser, 2016; Stevens et al., 2017); see also the discussion in Schwarz's Chapter 6 on presuppositions, in this volume. In conclusion, as these few examples already show, experimental investigations of prosody are essential to developing a comprehensive perspective on natural language meaning.

ACKNOWLEDGMENTS