

Focus perception

Looking at some of the properties of focus perception as observed by Krahmer & Swerts (2008).

1 Wait, what?

“Many linguists approach intonational matters from a purely speaker-oriented perspective. For instance, in different studies, in as far as these are empirical in nature, evidence for particular tonal distinctions is often solely based on acoustic analyses of fundamental frequency (F0) traces. However, if one wants to gain full insight into how intonation ‘functions’, such an approach is arguably incomplete.”

Ok, I see. Many who sew avoid bodily harm by wearing a thimble. For instance, different people who sew, insofar as they are not just pretending to sew, use metal thimbles. However, if one wants to succeed in ‘avoiding bodily harm’, the protection afforded by wearing a thimble is arguably incomplete. As anyone who has tried this for playing football will readily confirm.

Look. You can sew quite safely without needing to don full linebacker gear. The methods, goals, and hazards of sewing differ from those of football playing. Feel free to study intonation from a purely speaker-oriented perspective, that’s one thing you can do, and in so doing, you can sensibly rely on pitch tracks. So long as you aren’t then claiming that the world consists solely of spoken monologues to nobody (and thus, that, by understanding intonation production, you will have understood everything), there’s no problem here. Moreover, it’s fairly foolhardy to set your goal to be the acquisition of “full insight” into how intonation ‘functions’—unless you define “functions” and “full” in some way, this just winds up being the goal of understanding everything perfectly and completely. So, I’ve soured on this article already a bit right off the bat, even if this is just introductory fluff.

“A prosodic feature can only be said to be communicatively relevant if it is not only encoded in the speech signal by a speaker, but if it also has an impact on how an utterance is processed by a listener. In other words, claims about important intonational categories and their respective meanings are somewhat premature if they are not backed up by results that show that these are also relevant at the receiving end of the communication chain.”

That is to say: The only things worth studying are those things we say are interesting. Nothing prosodic is interesting unless it can be used to communicate. So, stop studying production. All that matters is perception. If what you discover about pro-

duction is something we can learn from studying perception, great, you've discovered something true (but the wrong way, all you needed was to look at perception). If you "discover" something about production that we couldn't back up in perception, you're doing it wrong—whatever crazy thing it is that you're doing, what you've learned is not among the things that are important to know.

I guess I find this so irritating because I hold almost entirely the opposite view. What you know about how phonology and prosody works in your language does not change depending on whether you talking to your friends or to your microwave. My categorization of /t/ and /d/ as different phonemes is a part of my language knowledge, and no less real just because the microwave can't distinguish between me saying "ten" and me saying "den." It just blinks "12:00" at me regardless.

If you want to understand and explain perception, study perception. That's fine, but it's not explaining our language knowledge. It might be explaining something about how we can put that knowledge to use, or even maybe isolating special perception-specific language knowledge. In fact, to the extent that we systematically produce things that are not perceivable, that deepens the mystery considerably—how would children learn how to do that? So, ultimately, what K&S are studying has interesting implications, but it seems as if they are interesting for reasons other than what K&S thought.

"For instance, the difference between H* and L+H*, as defined in the ToBI framework, has been claimed to indicate semantically distinct categories... Yet these two intonational categories are often confused by labellers..."

Fortunately, K&S had a panel "independent intonation experts" to do the labeling in their experiment, who are presumably immune to this effect.

- | | |
|-----------------------------------|---------|
| (1) TOTTENHAM ONE – LIVERPOOL one | English |
| (2) UDINESE UNO – ROMA UNO | Italian |

Italian strongly disfavors deaccentuation within NPs "or other syntactic constituents" (kind of broad, no?). Also, Italian is "non-plastic" (term taken from Vallduví (1992)), taken here to mean that information structure has a more significant effect on the word order than it does on the intonational tune.

2 Experimental materials

The utterances to be used were obtained from speakers ahead of time. Two speakers, A and B, are separated by a screen. Each has eight cards, each card shows a figure in a particular color. Four are in a stack in front of them, four are laid out in front of them. A's stack corresponds to B's row, and vice versa. Each also has a 8-member "list." A describes the card at the top of his/her stack, and puts it in the next slot on the list, while

B takes the described card from his/her row and puts it in the next slot on the list. The game ends when both players are out of cards (at which point, I guess, both have lost the game). Speakers generally found it to be a very easy game to play.

This sets up different discourse contexts. Each card has two properties. Each property could have been mentioned in the previous turn or not (or not at all, at the beginning of the game). So, the first utterance is “new” in both properties. After that, a property is “contrastive” if the previous utterance differed in that property, and “given” if the previous utterance was the same in that property. In Dutch, the first word is the adjective. In Italian, the second word is the adjective.

(3)	a.	blue square	NN	2-1
	b.	red circle	CC	2-1
	c.	yellow circle	CG	3-1
	d.	yellow triangle	GC	1-3

And, fair enough, if you read this list, you’ll notice that you’re reacting to the previous utterance. I used numbers above to kind of estimate what I felt my own intuitions about my pitches were as I read the list.

Other materials included: 8 Dutch speakers and 8 Italian speakers to create the stimuli, and 25 Dutch speakers and 25 Italian speakers to listen to them, and 25 more Dutch speaker to watch animated talking heads at the same time.

A three-person panel of “independent intonation experts” were given the (Dutch) stimuli to label them. Mostly, properties in focus (C) have a pitch accent, but half the speakers always end on a low tone (a low boundary tone “L%”), and the other half always end on a high tone (a high boundary tone, “H%”).

3 Dutch study (study 1)

The task was this: Listen to a given instance of *blauw vierkant* ‘blue square’ and decide which of the following was the preceding utterance:

(4)	a.	<i>rood vierkant</i> ‘red square’	CG
	b.	<i>blauwe driehoek</i> ‘blue triangle’	GC
	c.	<i>rode driehoek</i> ‘red triangle’	CC

Two random orders, no feedback on correctness, self-paced, no communication with experimenters.

Big numbers are on the diagonal. Mainly, they get it right. Least confusion on CG, most confusion on CC.

		Classified as			Total
		CC	GC	CG	
Context	CC	95	83	22	200
	GC	60	119	21	200
	CG	10	6	184	200

Table 1: Study 1 results (Dutch)

		Classified as			Total
		CC	GC	CG	
Context	CC	52	70	78	200
	GC	53	82	65	200
	CG	61	73	66	200

Table 2: Study 2 results (Italian)

For the different groups of speakers (high and low boundaries): Low boundaries caused CC to be “predominantly” classified as CC. Not so for high boundaries, whose CC are very often classified as GC. This is supposed to account for the high rate of misclassification of CC and GC.

4 Italian study (Study 2)

Same deal here—except that for the Italian case, the panel of experts unanimously agreed that both words in all contexts get a pitch accent. All have the same intonational contour (a flat hat, with a downstep). They all sound the same. So can Italian listeners tell the preceding contexts apart? Well, of course not. They essentially guessed.

- (5) What utterance did this instance of *triangolo nero* ‘black triangle’ follow?
- a. *rettangolo nero* ‘black rectangle’ CG
 - b. *triangolo viola* ‘violet triangle’ GC
 - c. *rettangolo viola* ‘violet rectangle’ CC

5 Study 3 (eyebrows)

Animated a head along with four of the voices collected from real people, plus two synthesized voices.

		Classified as			Total
		CC	GC	CG	
Context	$\hat{C}C$	64	41	45	150
	$C\hat{C}$	59	70	21	150
	$\hat{G}C$	34	91	25	150
	$G\hat{C}$	33	90	27	150
	$\hat{C}G$	16	22	112	150
	$C\hat{G}$	16	30	104	150

Table 3: Study 3 results (Dutch talking heads)

Same basic experiment, except there are eyebrows to watch. The eyebrow move is on either the first or second word, and—crossed with the three contexts—this results in six context-eyebrow configurations.

Where there was a single contrast (GC and CG), the subjects generally went with the prosody, no matter what the eyebrows were doing.

There was a bit of an effect of eyebrow movement for the CC cases. People generally had about a 40% chance of classifying this right no matter what the eyebrows were doing. (However, the fact that it is so low is possibly attributable to the fact that the eyebrows were doing *something*.) But when they classified it wrong, eyebrows on the second word (the noun) pushed them to overwhelmingly pick the noun contrast case, while eyebrows on the first word were essentially ignored.

That is, the only time the eyebrows mattered is if they were on the noun, when there were no prosodic cues.

6 Discussion and followup

What have we done? We have gained insight into “how useful pitch accents are as signals of focus.” Mm. And why is that useful?

7 Second-occurrence focus

Let me start with Beaver et al. (2007) and work my way from there.

The problem: Intonation can have a truth conditional effect (everyone got this now?)

- (6) a. Jan only gave Bill MONEY.
- b. Jan only gave BILL money.

The standard approach to this is that there is a “deep focus” of some kind on *money* in (6a) and on *Bill* in (6b). That is, in the syntactic construction of the sentence, the element in focus is marked in some way (i.e. a [+focus] feature is attached to it), which results in certain effects on the meaning, certain effects on the phonology, and possibly certain effect on the syntax (forcing the focus to move around, say).

It is usually assumed that the mappings are simple—if something is marked with [+focus], then it’s going to get something like a pitch accent in the phonology.

There is a set of things (such as *only* or *even* in English) that cause the interpretation to differ (truth conditionally) depending on the placement of focus. They are the FOCUS-SENSITIVE ITEMS (“operators”).

The meaning of *only*, for example, goes something like this: Presume that there is a set of alternatives, which are formed by substituting other things in where the focused part is. There are at least two such alternatives. The one that is actually in the sentence causes the sentence to be true. Any of the others in place of that will cause the sentence to be false.

So, the interpretation of a sentence with *only* crucially relies on the placement of focus to determine the alternatives.

Even works in a similar way: Presume that there is a nontrivial set of alternatives defined by the focus, and that the sentences that one forms by substituting in each of these alternatives can be ordered by likelihood. The alternative that is actually in the sentence is the least likely of the bunch, and the sentence is true substituting that one in. (Thus, you can conclude that it is almost certainly true for any of the other alternatives, since they are more likely).

It’s a bit complicated, but here are some examples to help us out.

- (7) a. Even PAT passed the test.
- b. I didn’t see even ONE policeman.
- (8) a. # Even EINSTEIN passed the test.
- b. # I didn’t see even FORTY-FIVE policemen.
- (9) a. Pat even gave BILL money. *Pat gave nobody else money.
- b. Pat even gave Bill MONEY. *Pat gave Bill nothing else.

The interpretations of each of these relies on the position of focus. It’s just what *only* and *even* mean. Which is why second-occurrence focus is kind of a puzzle.

- (10) 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.

The meaning is such that the focus *has* to be there in on *vegetables* in some sense. Otherwise it wouldn't mean what it does. But the pronunciation of *vegetables* sounds pretty normal in the second sentence, it doesn't sound like there's a focus on it.

So, is the phonology just doing its own thing? Why doesn't the [+focus] mark on *vegetables* result in a pitch accent there? Is this a "phonologically invisible focus"? If so, are we in danger of losing a grip on what focus is supposed to mean?

Beaver et al. (2007) set out to investigate experimentally exactly what is going on there with *vegetables* in the second sentence. The pitch does not appear to be going up. Or is it? Or is there something else that happens there? Can instruments detect it? Can *people* detect it?

People have usually assumed this as being a connection between semantic focus and phonological focus.

(11) FOCUS PROJECTION

An F-marked constituent must contain at least one phonologically focus-marked subpart.

8 Their experiment

We'll look at it more detail next time. It's alluded to by Büring (2008).

Basic conclusion: no pitch marking on SOF, but there was a small but significant lengthening and increase in energy. Perception study: hearers found the focus above chance (63%).

References

- Beaver, David, Brady Zack Clark, Edward Flemming, T. Florian Jaeger & Maria Wolters. 2007. When semantics meets phonetics: Acoustical studies of second-occurrence focus. *Language* 83(2): 245–276.
- Büring, Daniel. 2008. Been there, marked that: A theory of second-occurrence focus. Ms., UCLA.
- Krahmer, Emiel & Marc Swerts. 2008. Perceiving focus. In Chungmin Lee, Matthew Gordon & Daniel Büring (eds.) *Topic and focus: Cross-linguistic perspectives on meaning and intonation*, 121–137. Dordrecht: Springer.
- Vallduví, Enric. 1992. *The informational component*. Ph.D. thesis, University of Pennsylvania, Philadelphia.