Prosodic marking of information status in Dutch and Italian

Swerts, M.G.J.; Krahmer, E.J.; Avesani, C.

Published in:
Journal of Phonetics

Publication date:
2002

Journal of Phonetics (2002) **30,** 629–654
doi:10.1006/jpho.2002.0178
Available online at http://www.idealibrary.com on IDEAL®

Prosodic marking of information status in Dutch and Italian: a comparative analysis

Marc Swerts

IPO, Center for User–System Interaction, Eindhoven, The Netherlands and CNTS, Antwerp University, Belgium

Emiel Krahmer

BDM/Computational Linguistics, Faculty of Arts, Tilburg University, The Netherlands

Cinzia Avesani

Istituto di Fonetica e Dialettologia, CNR, Padova, Italy

Received 23rd December 1999, and accepted 9th April 2002

This article reports on a comparative analysis of accentuation strategies within Italian and Dutch noun phrases (NPs). Its goal is not only to gain insight into what speakers do, but also into how listeners' perception and interpretation of incoming speech in different languages is affected by the distribution of accents. To this end, use is made of a particular experimental paradigm, which makes it possible to compare accent patterns in different languages from an acoustic, perceptual and functional point of view. Accent patterns were obtained via a simple dialogue game played by eight Dutch speakers and eight Italian ones. In this way, target descriptions of all speakers were obtained in the following four contexts: all new, single contrast in the adjective, single contrast in the noun, and double contrast. The accent patterns in these Dutch and Italian utterances were then compared in three different studies. Study 1 looks at accent distribution and finds that, in Dutch, new and contrastive information are accented, while given information is not; in Italian, distribution is not a significant factor in distinguishing information status, since within the elicited NPs both adjective and noun are always accented, irrespective of the status of the discourse context. Study 2 consists of prominence tests to investigate whether the accents differ in the degree of perceived emphasis. In Dutch, information status is reflected in these prominence differences: single contrastive accents are perceived to be the most emphatic, and given words the least emphatic. In Italian, it is less clear how gradient differences between accents can be linked to aspects of the discourse context. Study 3 presents a functional analysis of accent patterns exploring whether listeners are able to reconstruct a preceding utterance on the basis of prosodic properties of the current utterance.

Address correspondence to M. Swerts, IPO, Center for User–System Interaction, Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands. E-mail: m.g.j.swerts@tue.nl.

0095–4470/02/$ - see front matter

While this is possible for Dutch listeners, this is not at all the case for Italian listeners.

1. Introduction

This article reports on a comparative analysis of accentuation strategies in Italian and Dutch to test the hypothesis (based on informal claims in Ladd, 1996) that these two languages differ in their prosodic marking of information status inside noun phrases (NPs). In Dutch (and in most other Germanic languages), pitch accents are used as pointers to new information or as signals of a contrast-relation (where there is some controversy whether or not these two kinds of accents are intonationally distinct, see Krahmer & Swerts (1998, 2001) for discussion). Deaccenting (defined as the absence of an accent on a word that might otherwise be expected to be accented) may be used to signal that the word refers to "given" information or information that can otherwise be expected in the discourse. Evidence for a close correspondence between information status and (de-)accentuation in Dutch comes from acoustic studies on distribution of accents in speakers' utterances (Terken, 1984; see also Hirschberg, 1993, for English), from perceptual experiments in which subjects are instructed to rate the acceptability of utterances with different accentuation patterns (Nooteboom & Kruyt, 1987), and from psycholinguistic research that investigates to what extent sentence processing is facilitated by accents on new information and absence of accents on given information (Terken & Nooteboom, 1987, van Donselaar, 1995; see also Cutler, 1984, for English).

These findings for Dutch and English are compatible with the idea of Chafe (1974) that speakers only highlight information which should be at the center of the hearer's attention. Such a cognitive explanation may lead to the expectation that the tendency for given items to be deaccented is a prosodic universal, but there are different studies that show that this is definitely not the case. First, for English, Terken & Hirschberg (1994) report that the fact that an item is previously mentioned in the discourse does not itself suffice to trigger deaccentuation, and that grammatical function and surface position are also important factors: in their study of experimentally elicited referring expressions, they found that expressions for which the grammatical function in the current utterance and in the preceding context are different (e.g., change from subject to object) are likely to be accented. More generally, it is not clear that the effect of deaccentuation, as found in very controlled types of data, generalizes to more spontaneous dialogues. Bard & Aylett (1999), for instance, report that only 15% of (within dialogue) repeated lexical items in their corpus of spontaneous dialogues were deaccented. Part of the problem with the analysis of these kinds of data is that it is often not trivial to exactly define information status, and that some repeated materials strictly speaking may not represent "given" information. For instance, natural conversations abound with feedback cues in which a speaker repeats what the speaking partner just said in a previous turn, cases that one would normally not treat as "genuine" given information (see, e.g., the study by Shimojima, Katagiri, Koiso & Swerts, 2002).

Stronger evidence against deaccentuation as a universal phenomenon comes from studies of languages other than Dutch and English (see, e.g., Cruttenden 1993, 1997). For instance, in Japanese, accent is a lexical feature and speakers do not have the option of accenting or deaccenting words according to their information status. But counterevidence against deaccentuation as a universal phenomenon can also be found in languages where accent is not a lexical feature. For instance, certain Romance languages, like Catalan and Spanish, sometimes resist deaccentuation, in particular at the end of intonation groups. Experimental evidence for this was found in Cruttenden (1993), in which native speakers of eight languages were asked to produce particular setting–response pairs where the response involved a lexical item repeated from the setting. In general, *plastic languages*[1] (Vallduví, 1992), such as Dutch and English, tend to use intonation to mark information status; on the other hand, *nonplastic languages*, such as Catalan and Spanish, make less use of intonation to mark information status, but also heavily rely on word order variation. While these findings do not entail that deaccentuation is impossible in certain Romance languages and while not all Romance languages are equally nonplastic, they do show that under certain conditions deaccentuation is infelicitous.

Prosodic studies of Italian, another nonplastic language, complicate the picture of accentuation patterns in Romance languages. On the one hand, some studies suggest that Italian accentuation strategies are similar to those of languages such as Dutch and English. In two cross-linguistic studies, Avesani, Hirschberg & Prieto (1995) and Hirschberg & Avesani (1997) showed that Tuscan Italian patterns with both Spanish and English in deaccenting given clauses and full NPs. In particular, all the Italian subjects of those studies deaccented clauses and simple NPs in sentences where a wide scope reading of a quantifier caused a subordinate clause or a full NP to be out of focus. Additionally, D'Imperio (1997) found that deaccenting of given information occurs in Neapolitan Italian and Farnetani & Zmarich (1997) reported that postfocal words in setting–response pairs of northern Italian have a shorter duration and a flat or falling F_0. On the other hand, Avesani (1997) showed that in samples of Italian broadcast speech items that have been previously mentioned in the same discourse segment can be accented again[2] irrespective of grammatical function and surface position in the sentence. Similarly, while Ladd (1996, 177–178 & p.c.) acknowledges that Italian allows deaccentuation on a sentence level (e.g., repeated full NPs may be deaccented), he claims that Italian strongly disfavors deaccentuation *within* NPs or other syntactic constituents. That there are such syntactic restrictions appears from the following Italian example (after Ladd, 1996), compared with literal English and Dutch translations, which gives some cases of accent patterns in adjectival phrases:

Italian	Correre è come camminare in FRETTA, soltanto si deve andare molto più in FRETTA
Dutch	Hardlopen lijkt op lopen met HAAST, alleen heb je veel MEER haast

[1]This term refers to Vallduví's idea that the prosodic pattern in a "plastic" language is "moulded" to fit the information structure.

[2]We will avoid the term "reaccented" as it has been used by Cruttenden (1993), given that this implies an unfortunate Germanic point of view, i.e., a word would normally be deaccented in English or Dutch, but instead is accented again.

English Running is like walking in HASTE, only you have to go much MORE in haste

According to Ladd, a pattern where the second *fretta* would be deaccented and the accent would "move" to *più* (like in the Dutch and English translations) is rejected by Italian native speakers, because it only deaccents a part of the adverbial phrase *molto più in fretta*.[3] In other words: one can deduce from Ladd's claims that Italian differs from Dutch in that the former does not allow moving accents inside syntactic constituents in order to convey contextual givenness. If this is taken literally, it would imply that within Italian constituents givenness cannot be encoded prosodically, unless, of course, the given accents were intonationally different from the others (where a further question would be whether the group of "other" accents can be divided into distinctive "contrastive" and "newness" accents).

This article describes a comparative analysis of Dutch and Italian to gain further insight into accentual differences between plastic and nonplastic languages. For a variety of reasons, it is difficult to base typological comparisons on results described in the literature. First, the framework of various researchers in these studies is often quite different, leading to confusion about even basic concepts such as "accent", which makes it difficult to judge to what extent attested prosodic differences between languages are due to the characteristics of the language structure itself, or to the framework of the researchers that investigated it. Similarly, the definitions of discourse related concepts, e.g., what counts as "given", "new" or "contrastive" information, are variable, so that it is hard to see whether such phenomena have a different impact on accent patterns in different languages. In addition, previous studies differ in that they make use of different types of speech materials, with some people focusing on isolated read-aloud utterances in citation form and others making use of completely spontaneous conversations. Given that it is known that different genres can affect accent patterns in different ways, this makes the analysis considerably more difficult. Related to this point, previous studies, especially those that are listener-oriented, often make use of constructed speech materials, either read aloud fragments (e.g., Gussenhoven, 1983), sometimes produced by a reader experienced in producing specific intonation contours (e.g., Terken & Nooteboom, 1987), or sentences with synthetically generated intonation contours (Kruyt, 1985; Nooteboom & Kruyt, 1987). With these types of data, there is a danger that one tests cases that are not representative of what happens in naturally occurring utterances. Finally, the studies on accent in particular languages may also differ in that they focus on different levels of analysis, e.g., acoustic, perceptual or functional. As a matter of fact, we do not know of any study that makes typological comparisons of prosodic structure in various languages on more than one level. Therefore, the current study aims to put the investigation into a broader perspective, in that it not only wants to gain insight into what speakers do, but also into how the listeners' perception and interpretation of incoming speech in different languages is affected by the distribution of different types of accents.

[3]Ladd's example is a bit unfortunate, because while *molto PIÙ in fretta* would indeed be rejected by Italian subjects, an accent shift to *molto* as in *MOLTO più in fretta* would not, which weakens his explanation that the rejection of the first is due to the fact that this pattern "deaccents only part of the adverbial phrase".

To solve all the methodological problems described above, we have decided to develop and use a particular experimental paradigm, which makes it possible to make typological comparisons of accent patterns in different languages from an acoustic, perceptual and functional point of view. An important aim of the series of experiments described below is to use naturally elicited speech data, in production, perceptual and functional tests, whose accent patterns will not be manipulated. The current study compares accentuation patterns in Dutch and Italian in three different ways: (1) we look at distributional differences in NPs that are elicited using exactly the same experimental paradigm; (2) we look at gradient differences between accents to see whether informational distinctions are reflected in differences in emphasis; and (3) we ask to what extent listeners from both languages are able to reconstruct the dialogue history from accent patterns in the current utterance. These studies will show that Dutch and Italian are markedly different regarding the accent patterns in the elicited utterances, in terms of their distribution, in terms of perceived gradient differences, and in terms of their functional validity for reconstructing the prior dialogue context. In particular, while the analyses at these three different levels show that in Dutch there is a close connection between information status and accent patterns, there is no such correspondence in the Italian NPs.

In the next section, we first describe the experimental design to elicit accent patterns in both Dutch and Italian utterances. The following sections then describe analyses of these speech data and present results on accent distribution, gradient differences between accents and functional differences between accent patterns in the two languages. We end with a discussion and a conclusion.

2. Collection of speech materials

Accent patterns for both languages were obtained in a (semi-) spontaneous way via a simple dialogue game played by four Dutch pairs and four Italian pairs of subjects, thus giving eight speakers for each language. The game is essentially an alignment task of figures played by two subjects, call them A and B, who are separated from each other by a screen. Figs 1 and 2 visualize the experimental setup with a bird's-eye perspective on the starting situation of the game and the situation after the first turn in the game. In each game, both players have an identical set of eight cards at their disposal, each card showing a geometrical figure in a particular color. Four of these cards are put on a stack in front of them, the other four cards are in a row before them. The four cards in the *stack* of A are the same as the four cards in the *row* of B, and *vice versa*. Both players also get a paper on which the numbers 1–8 are printed. In the course of the game, both players need to fill the blanks behind each number, from 1 to 8, with the cards with geometrical figures. The ultimate goal of the game is that both players will have generated an identical ordered list of geometrical figures on this paper. The game consists of a series of turns in which one participant gives instructions to select a card with a particular geometrical figure and the other follows these instructions. In each consecutive turn, the participants switch roles so that the original instruction-giver becomes the instruction-follower, and the other way around. A turn always consists of two actions on the part of the instruction-giver, and two related actions on the part of the follower. In turn 1, the instruction-giver, say A, begins by

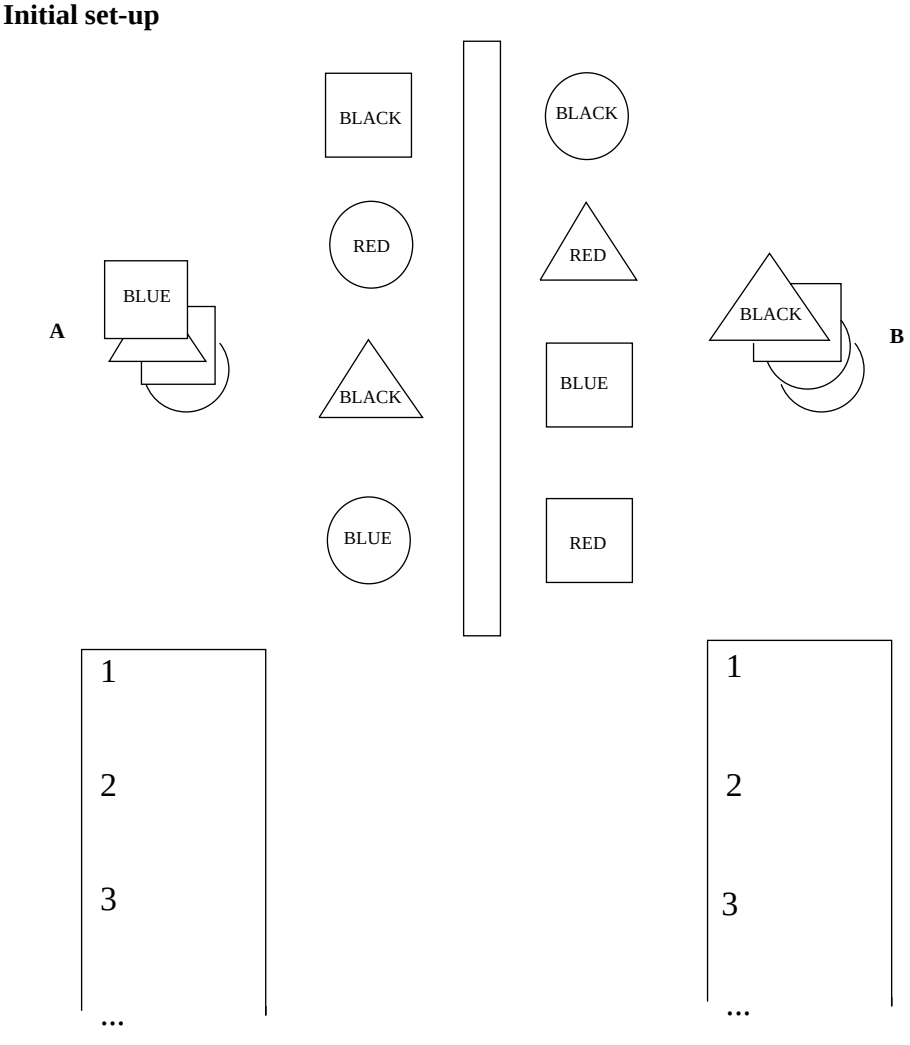

Figure 1. Visualization of the initial setup of the experiment to elicit different referring expressions (bird's-eye view). A and B represent the two participants in the dialogue game. In the actual experiment, the different figures were given different colors. Further explanations in the text.

describing the figure on the top of his stack (''a blue square''). After he has described this figure, he removes it from his stack and puts it behind number 1 on his list. The instruction-follower, B, listens to the description of A and removes that figure from his row of figures, and also puts it behind number 1 on his list. After the first turn of the game, the situation is now that both participants have the same figure behind number 1 on their respective lists: A has three figures left on his stack and has four figures in the row; B has four figures left on his stack, but has three figures left in his row. Now, the participants switch roles, so that B describes the figure that is on top of his stack (''a black triangle''), and A follows the instructions

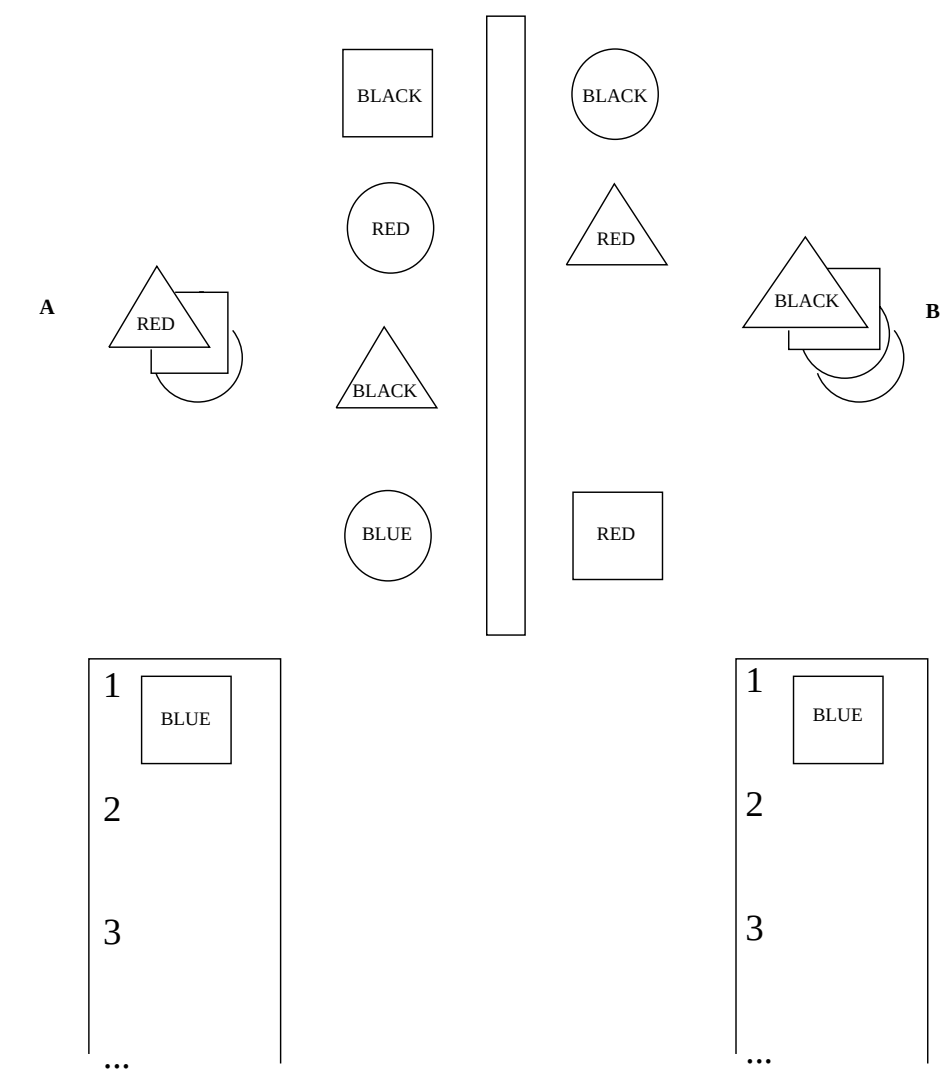

Figure 2. Visualization of the setup of the experiment after A's first move (bird's-eye view).

of B which will prompt both A and B to place the card with this object on the second place in the row with figures, and so on. The game is over when both players have no cards left. At that moment, they both have an identical list of eight figures on their respective papers. There are no winners or losers. After having completed a game, both participants have to start again with new cards, and this is repeated 8 times. There is always a break between two consecutive games of at least 2 min. This whole sequence of eight games is preceded by a short trial session. Note that the players are given the instruction to describe the figure on top of their stack in terms

TABLE I. Examples of the four contexts NN, CC, CG and GC

Context	Dutch	Italian
NN	(beginning of game)	(beginning of game)
	B: "blauwe vierkant"	B: "triangolo nero"
CC	A: "rode cirkel"	A: "rettangolo viola"
	B: "blauwe vierkant"	B: "triangolo nero"
CG	A: "rode vierkant"	A: "rettangolo nero"
	B: "blauwe vierkant"	B: "triangolo nero"
GC	A: "blauwe driehoek"	A: "triangolo viola"
	B: "blauwe vierkant"	B: "triangolo nero"

of its color and figure property, so they cannot add any particular discourse markers (e.g., the Dutch and Italian equivalent of "next"). In other words: speakers are allowed to only use noun phrases, so that each turn wholly consists of a particular phrase like "blauwe vierkant" in Dutch or "triangolo nero" in Italian. Speakers generally found it a very easy game to play, and as a consequence there are no faulty descriptions in the current data set.

The data thus obtained allow an unambiguous operationalization of the relevant contexts. A property is defined to be *new* (N) to the conversation if it is mentioned in the first turn of the current dialogue game, it is *given* (G) if it was mentioned in the previous turn, and finally a property is *contrastive* (C) if the object described in the previous turn had a different value for the relevant property. To avoid possible confusion, note that contrast defined in this way is something different from a *correction* (which one might interpret as a kind of meta-linguistic contrast). While corrections are also highly interesting from a prosodic/phonetic (not to mention a semantic) point of view, here we restrict ourselves to the narrower definition of (noncorrective) contrast. By varying the sequential order, target descriptions (Dutch: "blauw vierkant" (blue square); Italian: "triangolo nero" (black triangle)), which were the only phrases we used as data, were collected for eight Dutch and eight Italian speakers in four contexts: no contrast (all new, NN), contrast in the prefinal word (CG), contrast in the final word (GC), all contrast (CC).[4] Table I summarizes the situation. Note that in the Dutch elicited utterances the adjective always precedes the noun, whereas in the Italian data it follows the noun. In other words, if we refer to the first word in the elicited NPs, we mean the adjective in the case of the Dutch data, and we mean the noun in the case of the Italian data. All Dutch speakers were recruited from students and colleagues from IPO, speaking the variety of standard Dutch spoken in the Netherlands; the Italian speakers were all living in Italy, and were native speakers of the Tuscan variety of Italian.

The data obtained in this way from both Dutch and Italian speakers were used as input for typological comparisons of accent patterns in these two languages. The following sections will give results of these comparative analyses in terms of their distribution, in terms of gradient differences between accents and in terms of their functional validity.

[4]Note that in the two-letter abbreviations, the first letter corresponds to the contextual status of the first word, and the second letter to the contextual status of the second word.

3. Study 1: distributional differences

3.1. *Method*

All utterances of the target descriptions for the two languages (''blauwe vierkant'' (blue square) and ''triangolo nero'' (black triangle)) were used for a distributional analysis. For both the Dutch and the Italian data, the accent distributions were determined by the majority votes of three independent labelers for Dutch, and three independent ones for Italian. All labelers were intonation experts, and did not know the discourse context of the utterances while labeling the data. For Dutch, it was easy to get consensus on the accent specification, since there was full agreement between three labelers on accent distribution for 28 of the 32 utterances. The labeling of the Italian utterances was also easy, since there was consensus between the three labelers that all utterances have an accent on both the first and the second word.[5]

3.2. *Results and discussion*

The results of the distributional analysis are given in Table II. The Dutch data reveal a clear trend: in the NN (no contrast/all new) case, both adjective and noun are always accented. In other words, for the NN case, there is no ambiguity in the data between broad and narrow focus: the NN case always requires a double accent. When one item is given, while the other is contrasted (i.e., the CG and GC cases), the contrasted item is generally the only accented word and the given item is deaccented. Note that there is a complete lack of postnuclear accents in the CG case, while occasionally prenuclear accents on the adjective occur in the GC case. For the CC case, it appears that half of our speakers choose to accent both words, whereas half of the speakers only accent the noun. This last type of accent can be said to express broad focus, because a single accent on the noun indicates that the entire NP is contrastive (e.g., Ladd, 1996). Interestingly, the accent distributions can be related to differences between speakers, given that our Dutch speakers can be divided into two groups: four speakers who always end their utterance with a low-boundary tone, and four who consistently end it with a high-ending tone, producing a kind of ''list intonation''. It appears that our four high-ending speakers always provide CC cases with a single accent on the noun, whereas all utterances with double accent are produced by our low-ending speakers. This difference in tendency to have the first word accented or not may be due to the type of the final-boundary tone, and could be explained by the fact that speakers try to maximize the pitch difference between the onset and the closure of an intonation contour. The

[5]It is worth noting that the accents on the second word in the Italian phrase (the adjective) clearly are less prominent, which makes the accent less conspicuous than the one on the first word in a phrase, the noun. This seems to be due to the fact that the second word is, in general, clearly downstepped with respect to the first word. When we checked our majority labeling with two other phoneticians, they would sometimes classify the second word as deaccented. Yet, when we then confronted them with a realization comparable to a sequence of an accented and deaccented word as we would find in the Dutch data (which sounds rather marked in Italian, only used in special situations, e.g., to signal a meta-linguistic contrast), then they agree that there is a slight accent on the original utterances. (For a discussion of these issues, see Ladd, 1996; Section 5.2.1.3.) Therefore, we will stick to the consensus labels of the three labelers. In any case, the interpretation of the Italian results would remain essentially the same (there is no effect of discourse context on accent distribution), even if we relabeled some of the accented words as deaccented.

TABLE II. Accent distribution for the Dutch (NL) and Italian (I) target utterances in contexts NN, CC, CG and GC (see text). Accented words are printed in uppercase. The maximum number per condition is 8 (= the number of speakers)

Context	Dutch		Italian	
NN	Preceding: beginning of game		Preceding: beginning of game	
	BLAUWE VIERKANT	8	TRIANGOLO NERO	8
	BLAUWE vierkant	0	TRIANGOLO nero	0
	blauwe VIERKANT	0	triangolo NERO	0
CC	Preceding: rode cirkel		Preceding: rettangolo viola	
	BLAUWE VIERKANT	4	TRIANGOLO NERO	8
	BLAUWE vierkant	0	TRIANGOLO nero	0
	blauwe VIERKANT	4	triangolo NERO	0
CG	Preceding: gele vierkant		Preceding: rettangolo nero	
	BLAUWE VIERKANT	0	TRIANGOLO NERO	8
	BLAUWE vierkant	8	TRIANGOLO nero	0
	blauwe VIERKANT	0	triangolo NERO	0
GC	Preceding: blauwe driehoek		Preceding: triangolo viola	
	BLAUWE VIERKANT	2	TRIANGOLO NERO	8
	BLAUWE vierkant	0	TRIANGOLO nero	0
	blauwe VIERKANT	6	triangolo NERO	0

hypothesis would be that, in the case of low-boundary tones, speakers tend to produce an initial accent to have a high onset which makes a clear melodic contrast with the final-boundary tone, whereas the absence of an accent results in a low onset that gives a clear melodic difference with the final high.[6] Related to this, our data suggest some effect of speaker convergence because our speakers tend to copy the intonational strategies from their respective speaking partners: the high-ending speakers only communicate with other high-ending speakers, and the same is true for the low-ending speakers; so there are no cases where a low-ending speaker interacts with a high-ending speaker. Given the considerable differences in terms of the overall intonation contour produced by these two kinds of speakers, we will always give separate results for these two Dutch speaker types in the remainder of this article.

The Italian data are quite different in that every word is always accented, irrespective of context. In particular, given information is not deaccented, as in Dutch, but accented. All speakers produce the same kind of contour in all contexts, i.e., a flat hat shape with the second accent downstepped with respect to the first accent. Thus, it appears that accent distribution within NPs does not provide clues about information status in Italian, while it does in Dutch.

To provide an idea of the kinds of intonation patterns the speakers produced, Figs 3–5 give the F_0 contours for utterances elicited in the NN and CG context, for a low-ending Dutch speaker (JR), a high-ending Dutch speaker (WY) and an Italian speaker (AG). The figures show that, for the Dutch speakers, the contours elicited in

[6]A simpler explanation is that the definition of "accent" in a prenuclear context is unclear enough that the judges, not the speakers, are less inclined to mark a prenuclear accent in the high-ending than in the low-ending cases.

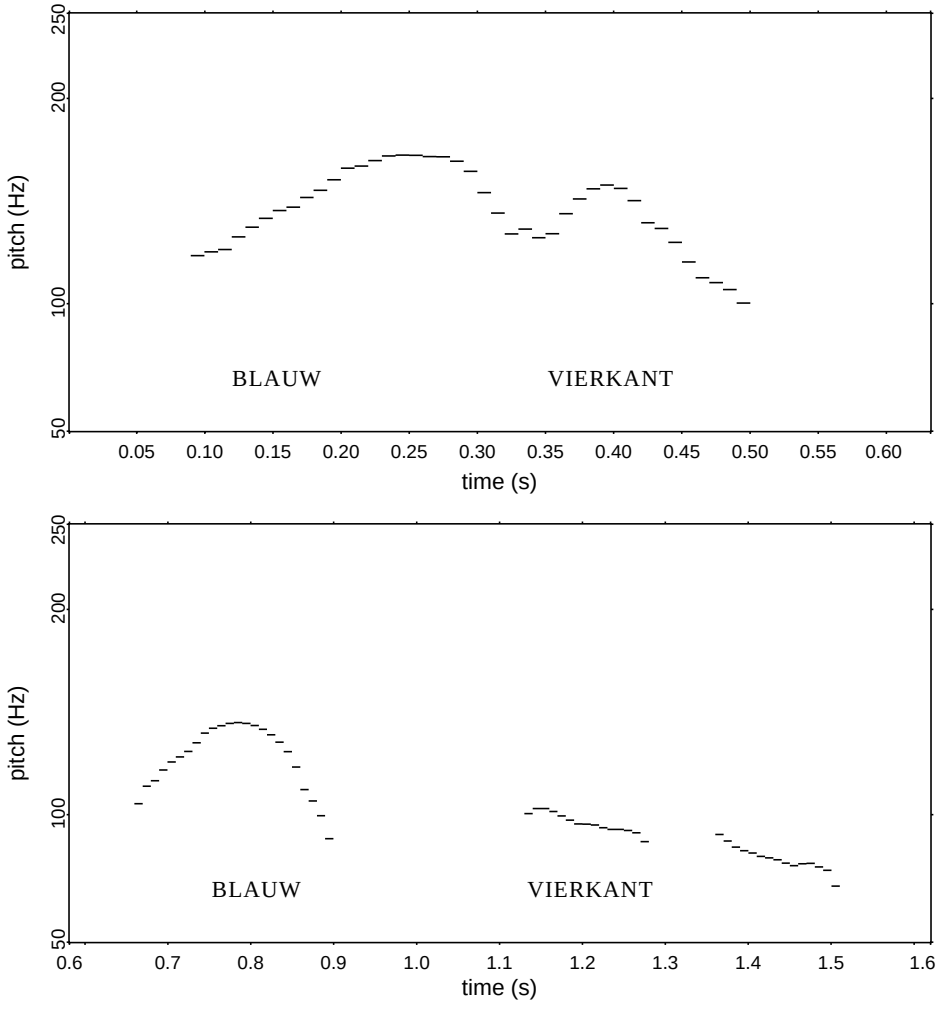

Figure 3. F_0 measurements of two realizations of "blauw vierkant" by speaker JR (low-ending Dutch speaker). Top: NN context. Bottom: CG context.

the NN and the CG context are clearly different. As can be seen in the bottom panels in Figs 3 and 4, the nuclear accent (the early timed fall for the low-ending speaker, and the rise followed by an additional final boundary-marking tone for the high-ending speakers) does not invariably fall on the last word in the utterance. More specifically, if the final word is deaccented (due to givenness) the nuclear accent shifts to the prefinal word (the adjective). As a result, for both Dutch speakers, the intonational pattern on the second word in the NN context is phonologically the same as the pitch movement that occurs on the first word in the CG context. In Krahmer & Swerts (2001), we have shown that it is precisely the presence of such a nuclear accent in a nondefault position which leads to a contrastive interpretation of these utterances. By contrast, in the Italian data, each

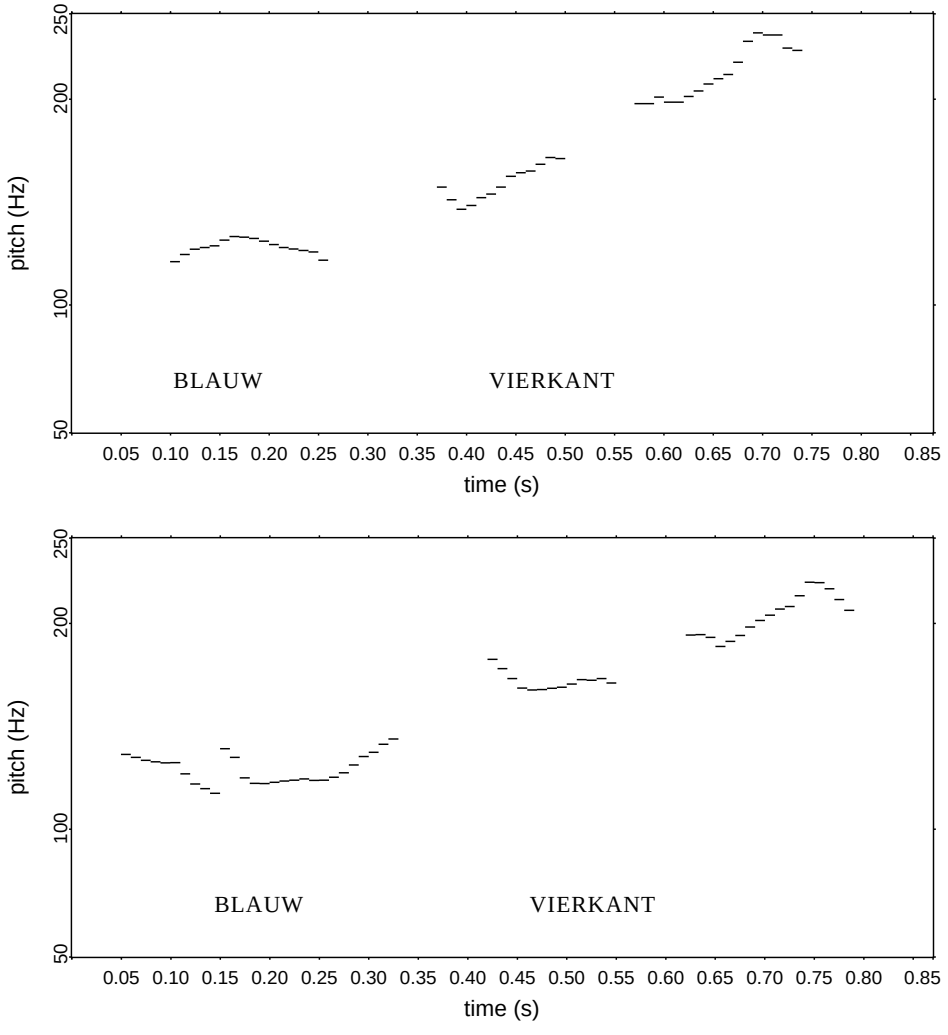

Figure 4. F_0 measurements of two realizations of "blauw vierkant" by speaker WY (high-ending Dutch speaker). Top: NN context. Bottom: CG context.

word is always accented, and thus our Italian speakers cannot move the accent from a default to a nondefault position. Consequently, as can be seen in Fig. 5, the F_0 patterns for "triangolo nero" are essentially the same for the two conditions, NN (top) and CG (bottom). Summarizing, while the discourse context clearly has an impact on accent distribution in the Dutch data, this is definitely not the case for the Italian utterances.

However, it might be that there still is an effect on the degree of prominence of the accents, where the hypothesis would be that given accents in Italian are less emphatic than the new or contrastive ones. For Dutch, it may be interesting to compare whether there are some gradient differences between newness accents and

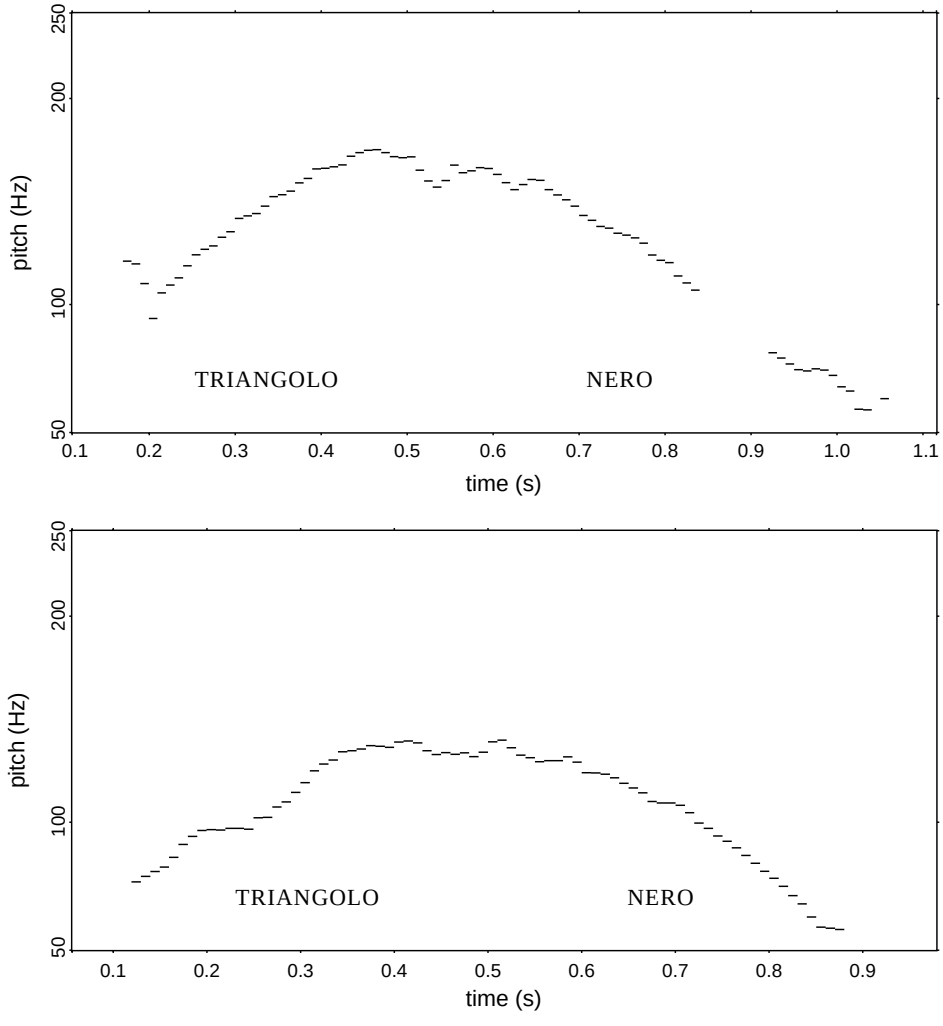

Figure 5. F_0 measurements of two realizations of "triangolo nero" by speaker AG (Italian speaker). Top: NN context. Bottom: CG context.

contrastive accents, given that it has been claimed before that there can be a difference of degree between these two accents (e.g., Hayes & Lahiri, 1991). Excursion size of movements has been mentioned in the literature as a good correlate of the degree of prominence of accents (e.g., 't Hart, Collier & Cohen, 1990). Therefore, we performed various acoustic analyses of excursion sizes, using the following procedure. All target utterances from all Dutch and Italian speakers were split into a first half containing the first word, and a second half with the second word. Then, for all these isolated words, we automatically determined the fundamental frequency F_0 by means of a method of subharmonic summation (Hermes, 1998), combined with dynamic programming to smooth the F_0 contour to get rid of any possible pitch errors. Then, the highest and lowest F_0 values were

automatically determined, which were then used to define the excursion size of a movement as the highest F_0 value minus the lowest F_0 value in a word. To get an idea of the relative accentual strength within a phrase, we looked at differences in excursion size between the first and the second word in a phrase. Accordingly, the following figures display the excursion size of the first word minus the excursion size of the second word for the four conditions. Fig. 6 summarizes the situation, which gives the results for all eight Italian speakers and those for the four low-ending and the four high-ending Dutch speakers separately.

While we have too few datapoints per condition to perform real statistical analyses, the observed patterns show some clear trends. First, the data in these figures support our earlier impressionistic observations on the intonational patterns observed in the Dutch and Italian data. It is true that for both the low- and the high-ending speakers, there is an effect of the discourse condition on the excursion sizes. In other words, while these two kinds of Dutch speakers provide their utterances with completely different intonation contours, their respective accent patterns are equally affected by discourse context. Concentrating on the pair CG and GC, one can see that the former has relatively high values for this measure on the first word, whereas in the GC condition the highest values occur on the second word; the data for NN and CC lie in between these two extremes. Note also that there is an overall difference between the low- and the high-ending speakers in that the former have smaller excursion sizes in the second word, while the reverse is true for the utterances of the high-ending speakers. For Italian, however, there is no effect of discourse context on the F_0 measures. Note also that, in the Italian data, the excursion sizes on the second word are consistently smaller than those on the first word, confirming our earlier observation that the second word in a phrase tends to be downstepped with respect to the first word in Italian.

4. Study 2: gradient differences

4.1. *Method*

The previous section showed that Dutch intonation is affected by information structure in a way that Italian is not, but the acoustic data may not tell the whole story since potentially there are other prosodic parameters that vary as a function of information context, e.g., duration, vocal effort, and other pitch correlates. Rather than exploring these parameters, we decided to set up a perception experiment to see whether listeners *perceive* any differences in prominence between the various accents. The goal of this experiment is to further explore whether Italian speakers, instead of marking context by means of the presence or absence of accents, do so by means of differences in degree of prominence. For Dutch, it is interesting to learn whether there are some emphatic differences between newness and contrastive accents.

Prominence ratings were obtained by means of a listening experiment (individual tasks) in which 16 Dutch and 16 Italian subjects (none of whom had served as speakers) participated. The Dutch participants came from various research institutes in the Netherlands and the Flemish part of Belgium. The Italian subjects were all from Tuscany. Each of them was presented with a series of pairs of phrases taken from the speech materials collected as described above: thus, they heard a pair containing an NN and a CG phrase, another pair with a CG and a GC phrase, etc.

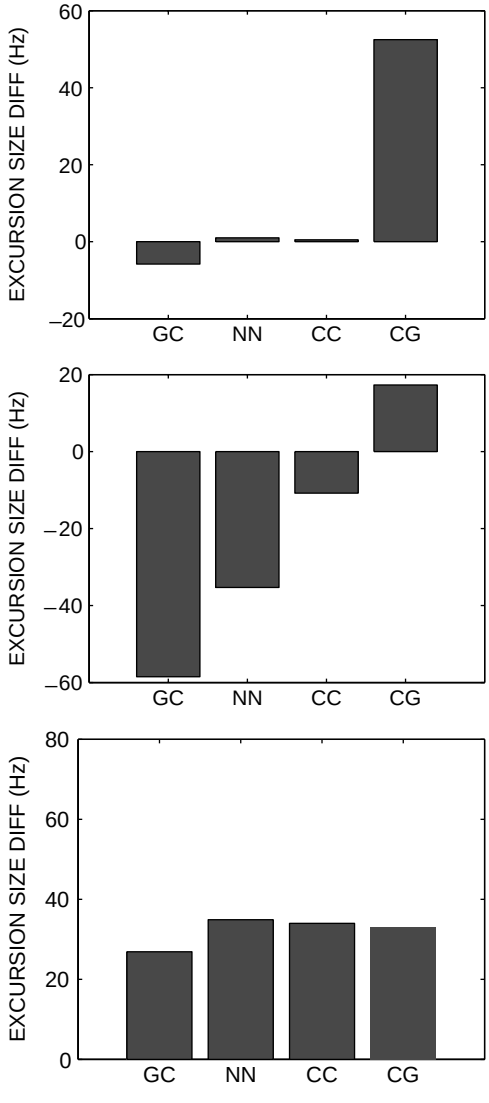

Figure 6. Average values in hertz (Hz) for differences between excursion sizes in first and second words for four Dutch low-ending speakers (top), four Dutch high-ending speakers (middle) and all eight Italian speakers (bottom) in four different contexts: CG, NN, CC and GC.

Stimuli were all realizations of ''blauw vierkant'' by two of our Dutch speakers (one using low-ending contours, the other using high-ending contours), and all realizations of ''triangolo nero'' of two Italian speakers. It was decided to use speech data from only two speakers for each language since performing all the pairwise comparisons with data from all eight speakers would lengthen the

experiment too much. The speakers were selected because they were representative in terms of accent distribution and had a clear voice and a fluent speaking style.

The procedure was as follows. Each utterance of a speaker was compared with all the other utterances of that speaker in two orders (AB and BA), giving 12 utterance pairs for each speaker of each language. Listeners had to do two tests. In one test, they were asked, for each pair of phrases, to attend to the first word (the adjective in the Dutch phrases, the noun in the Italian phrases), and were instructed to choose which of the two phrases had the most prominent first word; that is, subjects were comparing across phrases, not within phrases. In another test, using the same pairs of phrases, they had to pay attention to the second word (the noun in the Dutch phrases, the adjective in the Italian phrases), indicating which phrase contained the most prominent second word. The order of the two tests and the presentation of the pairs in each test were completely randomized. Listeners did not get feedback on the "correctness" or "incorrectness" of their responses. There was no further communication with the experimenters. No definition of prominence was given. Before they started the real experiment, subjects were asked to answer a few obligatory (name, region) and optional (contact address) questions, and had to do a brief training session (three pairs) to make them acquainted with the speech data and the experimental setting. Listeners could listen as often as they liked to each pair of utterances before giving a response. The experiment lasted between 10 and 15 min.

4.2. *Results and discussion*

Figs 7 and 8 show the results. The maximum value for each context is 96, i.e., 16 listeners $\times$ 3 contexts $\times$ 2 orders. In the following, capitalized letters indicate the words whose prominence the subject was rating (thus, 'Gc' indicated that the subject had to pay attention to the prominence of the prefinal word in GC condition). The results for Dutch show that words referring to given information (Gc or cG) uniformly score lowest on prominence (near zero). The no contrast/all new (nN/Nn) and double contrast (cC/Cc) cases basically score around chance level (50%). Single contrastive accents (Cg/gC) are always judged as the most prominent, for both speakers and for both the first and the second word. χ^2 tests revealed that all the different distributions for the Dutch data are significantly different from chance (JR, first word: $\chi^2 = 84.79$, df $= 3$, $p < 0.001$; JR, second word: $\chi^2 = 60.83$, df $= 3$, $p < 0.001$; WY, first word: $\chi^2 = 75.04$, df $= 3$, $p < 0.001$; WY, second word: $\chi^2 = 66.04$, df $= 3$, $p < 0.001$).

The picture for Italian is different (see Fig. 8). χ^2 tests show that all the different distributions for the Italian data are also significantly different from chance (CB, first word: $\chi^2 = 66.49$, df $= 3$, $p < 0.001$; CB, second word: $\chi^2 = 37.87$, df $= 3$, $p < 0.001$; AG, first word: $\chi^2 = 59.40$, df $= 3$, $p < 0.001$; AG, second word: $\chi^2 = 17.86$, df $= 3$, $p < 0.01$). Yet, it is harder to link these differences of degree to aspects of the discourse context. The results for the scores on the first word can be interpreted in terms of context, in that given words score lowest on prominence, while new words score highest. That these newness accents are rated more prominent than other accents may be due to a discourse effect: NN occurs only in the initial state of the game, which may give it extra emphasis (see Swerts & Geluykens, 1994). This is also clear from Fig. 5 which shows that the utterance in the NN context (top panel) is

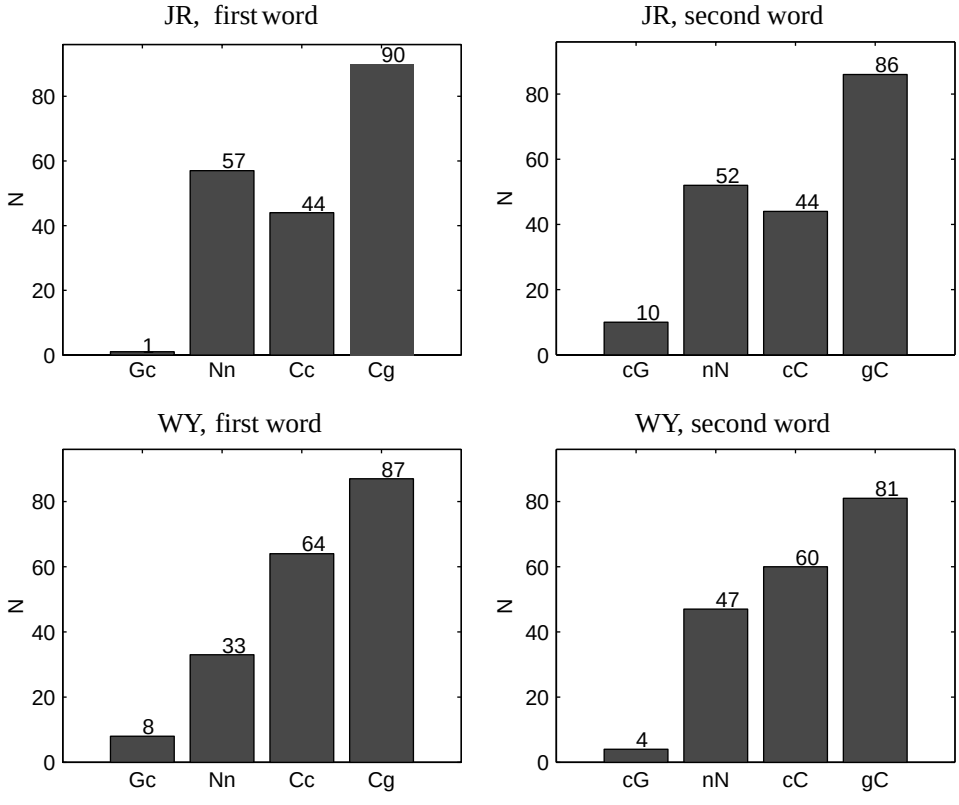

Figure 7. Prominence judgments (16 listeners) for first and second words for Dutch speakers JR and WY. The four contexts CG, NN, CC and GC are defined in the text; for readability the word of interest is capitalized. The maximum number on the *y*-axis is 96.

produced in a higher pitch range than the one in the CG context (bottom panel). Of course, for Dutch, NN also only occurs in the initial state but apparently single contrastive accents are more prominent than accents associated with the discourse effect.[7] However, the picture for the scores on the second word is less clear. Subjects indicated in comments after the test that they found the ratings of the second word far more difficult than those for the first word. The impression arises that the fact that the second word is realized with a downstepped contour leads to less accurate prominence ratings because of a "floor effect", i.e. due to a reduced pitch range on the second word, speakers have less variation space to signal subtle differences in prominence.

In other words: information status is clearly reflected in prominence ratings for the Dutch data (with low prominence for given words, and high prominence for

[7]Interestingly, single contrastive accents only stand out perceptually when entire utterances are presented to listeners; the effect disappears when listeners only hear single words, see Krahmer & Swerts (2001) for more details.

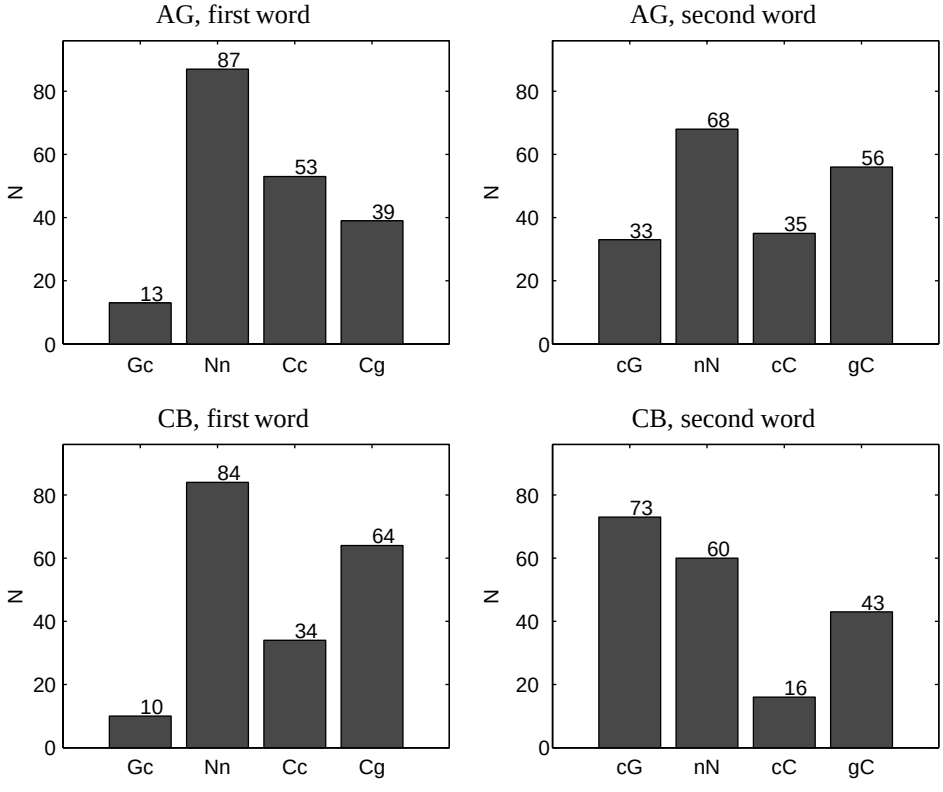

Figure 8. Prominence judgments (16 listeners) for first and second words for Italian speakers AG and CB. The four contexts CG, NN, CC and GC are defined in the text; for readability the word of interest is capitalized. The maximum number on the *y*-axis is 96.

single contrastive phrases). Such a correlation between prominence and information status is only true to a certain extent for the first word in a phrase in the Italian speech data, with "given" words being perceived as least prominent, and "new" words as most prominent. For the second word, no relation between perceived prominence and information status was found.

5. Study 3: functional differences

5.1. *Method*

From the results of a prominence experiment as described above, we do not know to what extent the accent structure of an utterance can be functionally relevant for listeners, for instance in that it steers the way they interpret the incoming utterances. Therefore, we set up a listening experiment using data from all speakers with the explicit aim of testing whether the accent structure of our target utterances had an impact on the way listeners process these utterances semantically. Given earlier

claims that speakers encode the prior discourse context in the accent structure of the current utterance, our general question was whether listeners are able to reconstruct the dialogue history from such prosodic cues. More specifically, they were instructed to guess, solely on the basis of a particular input utterance, what the information was in the utterance *preceding* that input utterance.

Dialogue reconstruction data were obtained from 25 native speakers of Dutch, and from 25 native speakers of Italian (mostly from Tuscany) for the Dutch and the Italian experiments, respectively. The experiment was performed on an individual basis and was self-paced. The stimuli used in the experiment were the Dutch and the Italian target utterances collected in the production experiment as described above. The Dutch subjects were presented with the different realizations of "blauw vierkant" (blue square) taken from their original context, and the task was to determine by forced choice what the preceding utterance was: (1) "rood vierkant" (red square), (2) "blauwe driehoek" (blue triangle) or (3) "rode driehoek" (red triangle). Similarly, the Italian subjects heard versions of "triangolo nero" (black triangle), and had to guess whether the preceding utterance was (1) "rettangolo nero" (black rectangle), (2) "triangolo viola" (violet triangle) or (3) "rettangolo viola" (violet rectangle). The corresponding contexts are CG (contrast in the first word), GC (contrast in the second word) and CC (all contrast), respectively.[8] All versions of the target utterance "blauw vierkant" and "triangolo nero" produced by the eight Dutch and the eight Italian speakers in the elicitation task were used, making a total of 24 stimuli (8 speakers $\times$ 3 contexts) for each language. Before the actual experiment started, subjects had to complete a few obligatory (name, region) and optional (contact address) questions, after which they entered a brief training session (three stimuli) to make them acquainted with the material and the setting of the experiment. No feedback was given on the correctness of their answers and there was again no communication with the experimenters. The entire experiment lasted between 5 and 10 min. Subjects could listen to each stimulus as often as they desired, although not much use was made of this option. The stimuli were presented in two different randomized lists, to compensate for potential learning effects.

5.2. *Results and discussion*

Fig. 9 contains the results for the data of all eight Dutch speakers taken together. The resulting overall distribution is significantly different from chance ($\chi^2 = 395.3$, $df = 4$, $p < 0.001$). The first thing to note is that for each context the bar with the "correct" classifications is the highest. This means that each context is most likely to be classified correctly. However, these chances are much higher in the case of single contrastive contexts (CG and GC) than in the double contrast case. Subjects are particularly good in reconstructing the dialogue history when the adjective is the single contrastive item (the classic case of narrow scope), which stands out prosodically due to the occurrence of a nuclear accent in a nonnuclear position. However, here too when it is the noun that is the single contrastive item, subjects are generally capable of reconstructing the context. Interestingly, the number of confusions with the double contrast context increases. This seems to imply that there

[8]Note that the all new situation (NN) was not incorporated in this experiment, because there are no utterances preceding the NN so that subjects cannot reconstruct the preceding utterance.

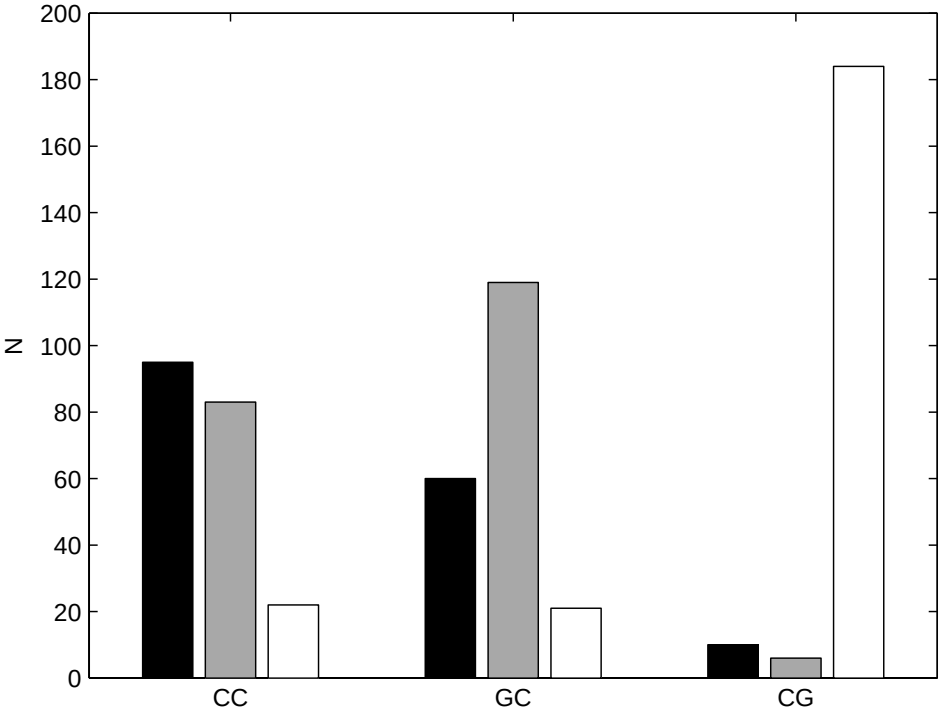

Figure 9. Classification of utterances as CC (black bar), GC (gray bar) or CG (white bar). Results for all eight Dutch speakers are taken together.

is at least some amount of broad focus/narrow focus ambiguity (but see below), although the narrow focus interpretation is still prevalent. This result is compatible with earlier findings from Gussenhoven (1983) and Rump & Collier (1996) that these ambiguous cases are more confusable than the CG case which only allows a narrow focus interpretation. In the case of double contrast, there appears to be a very substantial broad *vs.* narrow focus confusion. However, closer inspection of the data reveals an interesting difference between high- and low-boundary speakers.

Fig. 10 gives the results for each speaker separately, where each row in the figure shows data for speakers that formed a couple in the dialogue game (e.g., JR and RJ). The two top rows are low-ending speakers, the two bottom rows are high-ending speakers. For each of the speakers, χ^2 analyses reveal that the distribution is significantly different from chance. The results for these speakers are comparable in the sense that one always gets high classification scores for the CG cases. However, if we group the four low-ending speakers and the four high-ending speakers into two classes, we get a significant difference between these speaker types (Pearson $\chi^2 = 73.8$, df $= 8$, $p < 0.001$). The main difference between the two kinds of speakers is found for the double contrast (CC) case. For low-boundary speakers, utterances made in a CC context are predominantly classified as CC. Strikingly, this is not the case for high-ending speakers, whose CC utterances are very frequently classified as GC utterances, which matches our earlier observation that these speakers tend to

produce all-contrast utterances with a single accent on the noun. Thus, the fact that in Fig. 9 CC utterances are often misclassified as GC utterances is essentially due to the difference between low- and high-ending speakers rather than broad *vs.* narrow focus interpretations.

Figs 11 and 12 give the classification results for the Italian data, for all speakers taken together and for each speaker separately, with each row in Fig. 12 representing the data of speakers that formed a couple in our dialogue game (e.g., FB and PB). Neither the overall nor the speaker-specific results are significantly different from chance, except for the distribution of the scores for speaker PB. The overall results for the Italian data are significantly different from the overall results for the Dutch data (Pearson $\chi^2 = 223.8$, df = 8, $p < 0.001$). As one would expect on the basis of the earlier results on accent distribution and on prominence differences between accents, Italian listeners are never able to reconstruct the prior dialogue context on the basis of prosodic properties of the current utterance. Even for the one distribution, of speaker PB, that reaches significance, it is hard to see how the classification results can be related to information status.

6. General discussion

This article has focused on Italian and Dutch to see whether these languages differ regarding their use of prosody, in particular accent structure, to signal information status, i.e., whether information is given, new or contrastive. We have investigated this general question by applying a particular experimental approach, which consists of a dialogue game to elicit target utterances in different discourse contexts, and a series of studies that explore accent distribution, gradient differences between accents, and functional evaluation of accents. This paradigm is especially suitable for making crosslinguistic prosodic comparisons, and a variant of it has been applied in the meantime to Tokyo Japanese (Swerts, Taniguchi & Katagiri, 2000). Japanese is interesting from a typological point of view, given that accents are lexically determined in this language so that there may be other prosodic correlates of information status.

As to the results of the current study, we have found that the two languages, Dutch and Italian, are completely different regarding accent patterns inside NPs. In Dutch, it appears that accent patterns are indeed used to mark information status: accent distribution is the main discriminative factor with new and contrastive information generally accented, while given information is deaccented. In addition, the newness and contrastive accents appear to be phonetically different, in that the latter are perceived as more emphatic than the former (but see footnote 6). The third study shows that our Dutch listeners are capable, in the majority of the cases, to reconstruct the prior dialogue utterance on the basis of properties of the current utterance. Italian differs from Dutch in terms of accent structure at all three levels of analysis: distribution, emphasis and function. First, distribution is not a significant factor in this language, since within the elicited NPs both adjective and noun are always accented, regardless of the information status. There is some evidence that our speakers use gradient differences to distinguish given accents from newness accents, but this is only true for the first word in the NPs. As a result, it is

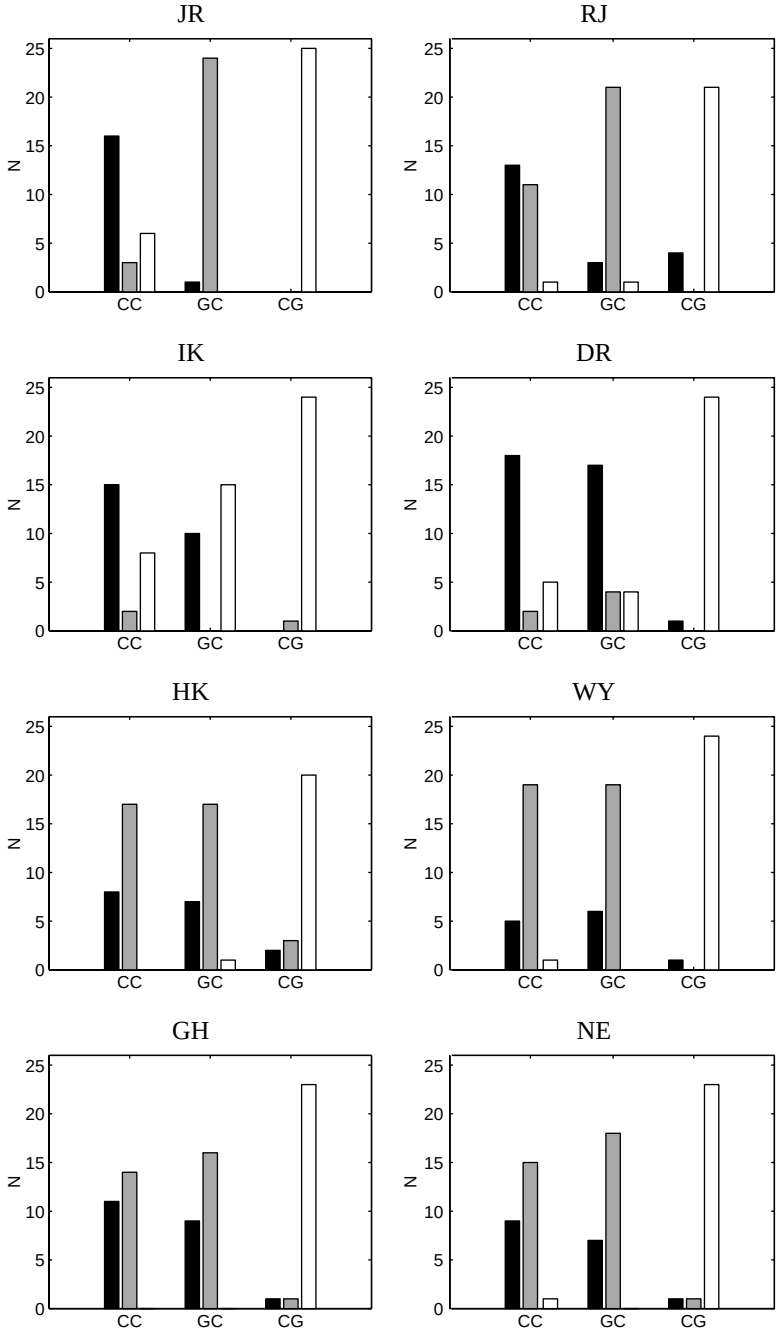

Figure 10. Classification results of utterances as CC (black bar), GC (gray bar) or CG (white bar). Results for each Dutch speaker are taken separately. Each row in the figure represents data of speakers that formed a couple in the dialogue game (e.g., JR and RJ).

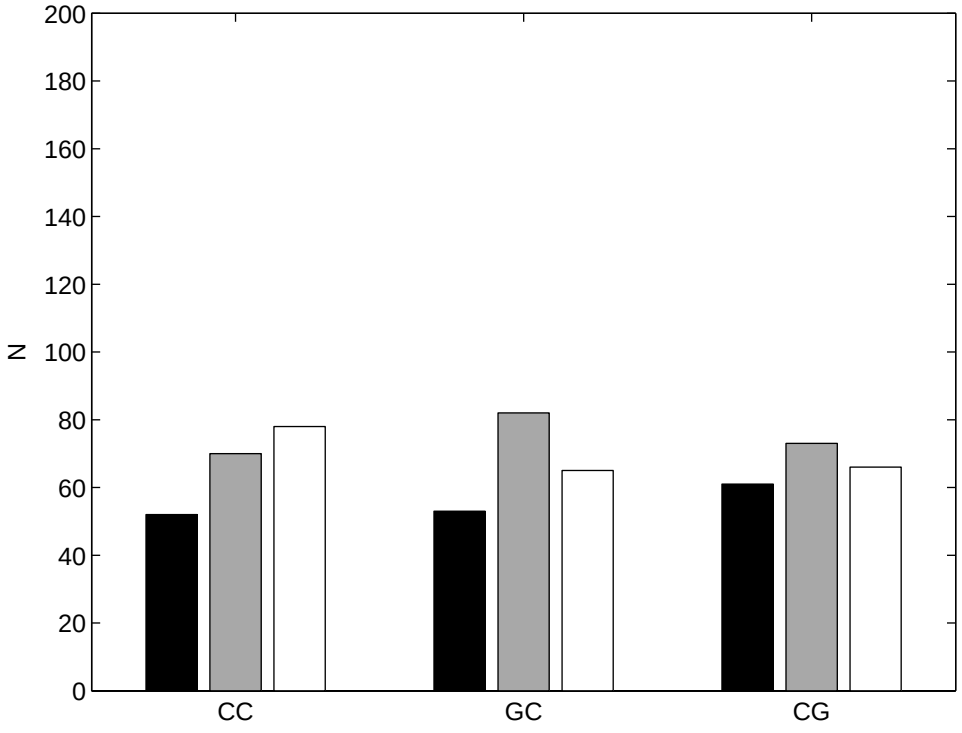

Figure 11. Classification results of utterances as CC (black bar), GC (gray bar) or CG (white bar). Results for all eight Italian speakers are taken together.

not surprising to see that the Italian listeners fail completely to interpret the target utterances in terms of the dialogue history.

It should be added that Italian, being a *nonplastic* language, has other means besides prosody of marking information status. For instance, it has a freer word-order than *plastic* languages such as Dutch, and it is known to exploit this freedom to mark information status. However, the constraints of the experimental paradigm did not offer any room for Italian speakers to use word-order as an indicator of information status. Therefore, it would be interesting to look for an experimental setup in which speakers have more freedom to describe a particular state of affairs. This might also shed a different light on the deaccentuation debate, given that Ladd claims that deaccentuation of *complete* NPs within a sentence is quite possible in languages like Italian, which is supported by data from previous studies (Avesani *et al*., 1995; D'Imperio, 1997; Hirschberg & Avesani, 1997). In addition, we have started to extend the experimental paradigm to also investigate multimodal cues to information status (Krahmer, Ruttkay, Swerts, & Wesselink, 2002): besides prosody, there are potentially other nonverbal cues, such as eyebrow movements or head nods, that may function as markers of discourse structure. To gain insight into the relative contributions of prosodic and visual cues to information status, we have started to redo the dialogue reconstruction experiment of Study 3, using a synthetic, talking face whose prosodic and visual parameters are systematically varied.

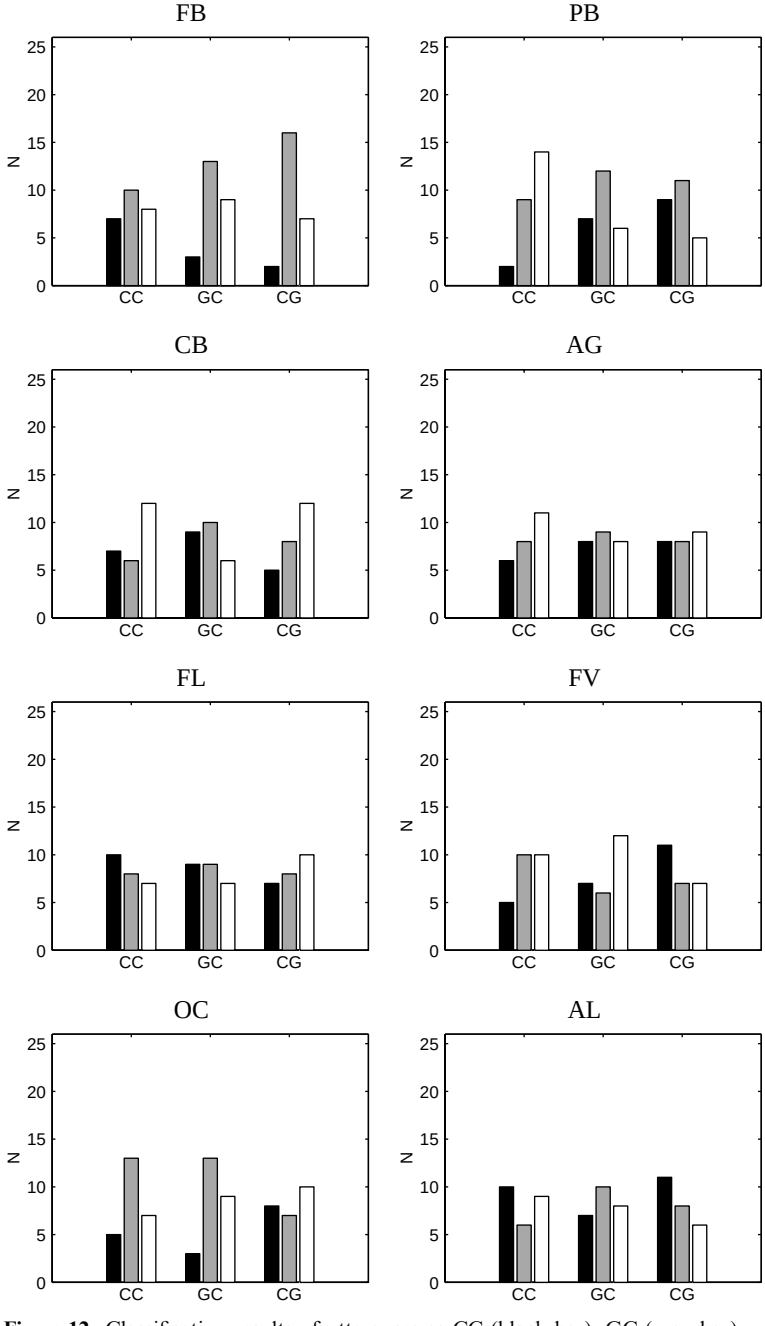

Figure 12. Classification results of utterances as CC (black bar), GC (gray bar) or CG (white bar). Results for each Italian speaker are taken separately. Each row in the figure represents data of speakers that formed a couple in the dialogue game (e.g., FB and PB).

Swerts is also affiliated with the Flemish Fund for Scientific Research (FWO – Flanders). The authors would like to thank Theo Veenker, Leo Vogten and Kees Kuijpers for technical assistance. Thanks to David Beaver, Bob Ladd, Gerry Docherty and two anonymous reviewers of Journal of Phonetics for discussion and various comments on an earlier version of this article, and to the 98 subjects (49 Dutch and 49 Italian) who volunteered to participate in our experiments as speaker or listener. Preliminary parts of this work have been presented at the following conferences and workshops: ICSLP 1998 (Sydney), ICPhS 1999 (San Francisco), the ESSLLI workshop on focus 1999 (Utrecht), the LSA workshop on topic and focus 2001 (Santa Barbara) and Eurospeech 2001 (Aalborg).

References

Avesani, C. (1997) I toni della RAI. Un esercizio di lettura intonativa. In Gli italiani trasmessi: la radio. Firenze: Accademia della Crusca, pp. 659–727.

Avesani, C., Hirschberg, J. & Prieto, P. (1995) The intonational disambiguation of potentially ambiguous utterances in English, Italian and Spanish. In *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, pp. 174–177.

Bard, E. G. & Aylett, M. P. (1999) The dissociation of deaccenting, given-ness, and syntactic role in spontaneous speech. In *Proceedings of the XIVth International Congress of Phonetic Sciences*, San Francisco, pp. 1753–1756.

Chafe, W. (1974) Language and consciousness, *Language*, **50,** 111–133.

Cruttenden, A. (1993). The de-accenting and re-accenting of repeated lexical items. In *Proceedings of the ESCA workshop on Prosody*, Lund, pp. 16–19.

Cruttenden, A. (1997) *Intonation*, 2nd edition. Cambridge: Cambridge University Press.

Cutler, A. (1984) Stress and accent in language production and understanding. In *Intonation, accent and rhythm. studies in discourse phonology* (D. Gibbon & H. Richter, editors) Berlin, New York: de Gruyter. pp. 77–90.

van Donselaar, W. (1995) *Effects of accentuation and given/new information on word processing*. Doctoral dissertation, Utrecht University.

Farnetani, E. & Zmarich, C. (1997) Prominence patterns in Italian: an analysis of F_0 and duration. In *Proceedings of the ESCA workshop on intonation*, Athens, pp. 115–118.

Gussenhoven, C. (1983). Testing the reality of focus domains, *Language and Speech*, **26,** 61–80.

't Hart, J., Collier, R. & Cohen, A. (1990) *A perceptual study of intonation. An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Hayes, B. & Lahiri, A. (1991) Bengali intonational phonology, *Natural Language and Linguistic Theory*, **9,** 47–96.

Hermes, D. J. (1988) Measurement of pitch by subharmonic summation, *Journal of the Acoustical Society of America*, **83,** 257–264.

Hirschberg, J. (1993) Pitch accents in context: predicting intonational prominence from text, *Artificial Intelligence*, **63,** 305–340.

Hirschberg, J. & Avesani, C. (1997) The role of prosody in disambiguating potentially ambiguous utterances in English and Italian. In *Proceedings of the ESCA workshop on intonation*, Athens, 1997, pp. 189–192.

D'Imperio, M. (1997) Narrow focus and focal accent in the neapolitan variety of Italian. In *Proceedings of the ESCA workshop on intonation*, Athens, pp. 87–90.

Krahmer, E. & Swerts, M. (1998). Reconciling two competing views on contrastiveness. In *Proceedings of the International Conference on Spoken Language Processing '98*, Sydney, 1998, pp. 1299–1302.

Krahmer, E. & Swerts, M. (2001) On the alleged existence of contrastive accents, *Speech Communication*, **34**(4), 391–405.

Krahmer, E., Ruttkay, Z., Swerts, M. & Wesselink, W. (2002) Pitch, eyebrows, and the perception of focus. *In Proceedings of the Speech Prosody 2002*, Aix-en-Provence, pp. 443–446.

Kruyt, J. G. (1985) *Accents from speakers to listeners. An experimental study of the production and perception of accent patterns in Dutch*. Doctoral dissertation, Leiden University.

Ladd, D. (1996) *Intonational phonology*. Cambridge: Cambridge University Press.

Nooteboom, S. G. & Kruyt, J. G. (1987) Accent, focus distribution and the perceived distribution of given and new information: an experiment, *Journal of the Acoustical Society of America*, **82,** 1512–1524.

Rump, H.H. & Collier, R. (1996) Focus conditions and the prominence of pitch-accented syllables, *Language and Speech*, **39,** 1–15.

Shimojima, A., Katagiri, Y., Koiso, H. & Swerts, M. (2002) Informational and dialogue-coordinating functions of prosodic features of Japanese echoic responses. *Speech Communication*, **36** (1/2), 113–132.

Swerts, M. & Geluykens, R. (1994) Prosody as a marker of information flow in spoken discourse, *Language and Speech*, **37,** 21–43.

Swerts, M., Taniguchi, M. & Katagiri, Y. (2000) Prosodic marking of information status in Tokyo Japanese. In: *Proceedings of the International Conference on Spoken Language Processing (ICSLP 2000)*, Beijing, pp. 254–257.
Terken, J. (1984) The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech*, **27,** 269–289.
Terken, J. & Hirschberg, J. (1994) Deaccentuation of words representing 'given' information: effects of persistence of grammatical function and surface position, *Language and Speech*, **37**(2), 125–145.
Terken, J. & Nooteboom, S. G. (1987) Opposite effects of accentuation and deaccentuation on verification latencies for 'given' and 'new' information, *Language and Cognitive Processes*, **2,** 145–163.
Vallduví, E. (1992) *The information component*. New York: Garland Press.