RESEARCH

Post-focal and factive deaccentuation in Persian

Hamed Rahmani, Toni Rietveld and Carlos Gussenhoven
Department of linguistics, Radboud University Nijmegen, 6500 HD Nijmegen, NL
Corresponding author: Hamed Rahmani (h.rahmani@let.ru.nl)

Two experiments were carried out to examine whether the Persian word accent is deleted in two putative deaccenting contexts, post-focal regions and presupposed embedded clauses, to the extent that accentual minimal pairs become homophonous. A production experiment showed low F0 plateaus on the post-focal and presupposed words, while a perception experiment showed that such words are not recognized above a just-noticeable-difference baseline. The results confirm that accents are deleted in the contexts concerned and that accent location contrasts are neutralized.

1 Introduction

A claim about Persian prosody that has frequently been made is that the word prominence disappears after the focus constituent (Eslami 2000; Vahidian-Kamyar 2001; Sadat-Tehrani 2007; Scarborough 2007; Hosseini 2014). To the best of our knowledge, no systematic experimental data are available to substantiate this position, while there has been a claim that pitch range compression after the focus does not neutralize contrastive positions of the word prominence (Abolhasanizadeh et al. 2012). The purpose of this contribution is to provide production and perception data as evidence that the word prominence is in fact removed after the focus as well as in certain presupposition constructions, to the extent that members of minimal pairs become homophonous. The paper is organized as follows. Subsections 1.1 and 1.2 give an overview of word accent and provide minimal accent pairs, respectively. Constructions that putatively trigger deaccenting are addressed in Subsections 1.3 (focus constructions) and 1.4 (presupposition constructions). Section 2 reports the design and the results of a production experiment, while Section 3 does the same for a perception experiment. In Sections 4 and 5, we present our discussion and conclusions, respectively.

1.1 Persian word accent

Although Persian word prominence has been described as "word stress" (e.g., Ferguson 1957; Windfuhr 1979; Mahootian 1997), recent experimental evidence shows that it is realized only through fundamental frequency (F0) (Abolhasanizadeh et al. 2012). Thus, in technical terms, it is an *accent*, a tone in a specific location, like the lexical (prosodic) prominence in Japanese and unlike the English *stress*, which is signalled by phonetic features other than just F0 (Beckman 1986). Throughout this text, we will use the term

accent for the Persian word prominence. Phonologically, the Persian accent consists of a H tone.[1]

Persian morphological words are accented on the final syllable without exception. By morphological word (henceforth Mword) we mean both morphologically simplex words (consisting of a simple stem) and complex words, i.e. derived words (consisting of a stem plus some derivational affix(es)) and compounds (consisting of more than one stem). The hallmark of Mwords is that they are treated as single units by inflectional and syntactic rules. Thus, for example, Persian compounds are indivisible constructions whose components cannot be interrupted by inflectional markers or modal particles.[2] Note that our definition of Mword constitutes only major word classes like noun and adjective. Some grammatical word classes will be briefly discussed in Subsection 1.2.

Three Mwords are illustrated in (1), where (1a) contains a simple stem, (1b) a stem plus a derivational suffix, and (1c) a stem-stem compound. Each example is given in four lines. The first line gives the IPA transcription with morpheme boundaries. Suffix boundaries and clitic boundaries are shown by "–" and "=" respectively, whereas components of a compound are separated by "+". The second line gives the prosodic representation. Phonological words (to be defined below) are marked by round brackets, while periods show syllable boundaries. Accent is marked with an acute over vowels. The third line gives the morpheme-by-morpheme gloss, when relevant. The fourth line gives the English translation.

(1) a. doxtar
         (dox.tár)
         girl

     b. doxtar-une
         (dox.ta.ru.né)
         girl-DER
         'girlish'

     c. doxtar+amme
         (dox.tar)(am.mé)
         girl+aunt
         'cousin'

The position of the accent in a Persian Mword always coincides with its right edge, regardless of the phonological word structure, with which Mwords are pervasively non-isomorphic (cf. Nespor & Vogel 1986). The most obvious cue to phonological wordhood in Persian is syllabification (cf. Bijankhan 2005; Hosseini 2014: 42). We therefore define the phonological word (henceforth Pword) as the domain of obligatory syllabification, meaning the Maximum Onset Principle (Kahn, 1976) is applicable across its entire domain, with a syllable structure (C)V(C)(C).[3] This formulation leads to the generalization that a Persian Pword corresponds to a simple stem plus its bound

---

[1] Persian has a small tonal inventory. In addition to the accent (H), there are two intonational boundary tones: L% for declaratives and Wh-questions, and H% for yes/no questions (cf. Mahjani 2003).

[2] For a similar approach to morphological wordhood in Persian, see Lazard (1992), Ghomeshi (1996) and Kahnemuyipour (2003).

[3] Hosseini (2014) has provided instrumental evidence which suggests that the phonological word domain derived by syllabification is also relevant for such prosodic effects as uninterruptibility and initial aspiration. First, it is unacceptable to put a pause in the middle of a Pword, and second, Pword-initial stops are significantly more aspirated than non-initial stops.

morphemes (an extension of this generalization will be given in Subsection 1.2, where we introduce *clitics*). Thus, while suffixes form a single Pword with their stem on the left, as illustrated in (1b), the constituents of a compound form separate Pwords, as shown in (1c). However, in all of (1a, b, c), there is only one accented syllable, the final syllable of the Mword.[4],[5]

The final accent pattern is not affected by the length and internal complexity of the Mword (cf. Vahidian-Kamyar 2001). Thus, the reduplicated compound [doxtar + amme + moxtar + ammé] 'cousin and stuff', a single Mword realized as four Pwords, receives only one accent on its final syllable.[6] Moreover, the alignment of accent to the right edge of Mword is fixed, meaning that it is not possible to accent the first constituent of a compound, say, not even if it is narrowly focused (see Subsection 1.3 for more discussion on this).

Before moving on, a remark should be made about two types of syllabification in Persian. Obligatory syllabification takes place in the Pword domain, as stated above. Optional resyllabification occurs (e.g., in fast speech) across Pwords, to the extent that all Pwords within a sentence-size utterance can be syllabified together (cf. Hosseini 2014: 43).[7] Thus, the compound in (1c) may be pronounced as [dox.ta.ram.mé]. Variations in syllabification in no way affect the position of accent. That is, a compound is always accented on the final syllable, regardless of its syllabification.

1.2 Minimal pairs

The location of accent has been generally recognized as highly contrastive in Persian, in that it contributes to distinguishing between morphosyntactically different utterances (Ferguson 1957; Lazard 1992; Vahidian-Kamyar 2001). An important source of these contrasts is a closed set of morphemes which are integrated into a single Pword with a morphological host on their left. While these cliticizing morphemes are similar to suffixes in their lack of prosodic autonomy, they differ from suffixes in that they are syntactically independent units (cf. Ghomeshi 1996; Kahnemuyipour 2003). In fact, they have a variety of functions that represent positions in the syntactic phrase, like auxiliaries, pronouns, conjunctions and focus governing morphemes. We refer to them as *clitics*, because they fulfil classic criteria for clitichood, such as not being particular about the word class of their host and failing to exhibit arbitrary gaps in their combinations with their host (Zwicky & Pullum 1983).

Crucially, since clitics, unlike suffixes, fall outside the morphological domain of the host, they are not assigned accent, causing the accent on the Mword-plus-clitic combinations to remain on the final syllable of the Mword. This property of clitics

---

[4] While for most compound structures, internal constituents are independent Pwords, exceptions to this generalization are not uncommon, in particular for frequently-used compounds with idiosyncratic meaning. For example, [gerd + ɒb] (round + water) 'whirlpool', is one Pword as evidenced by the fact that it is obligatorily syllabified as [ger.dɒ́b].

[5] The observation that Pwords in Persian are syllabification domains and that the constituents of compounds are assigned Pword status has been made for many languages like Greek, Turkish and German (Nespor & Vogel 1986).

[6] In this type of nominal reduplication, which is sometimes referred to as "echo reduplication", the first segment of the base (here the compound noun [doxtar + amme] 'cousin') is replaced by [m] or [p] in the reduplicant. Echo reduplications are used to express a wide range of meaning such as generality or multiplicity of entities. They are spontaneously and productively formed in everyday use. The idea that echo reduplications are single Mwords is supported by the fact that they are indivisible constructions and their syntactic behaviour is the same as that of the simplex nouns (Ghaniabadi et al. 2006).

[7] Hosseini (2014) describes this optional resyllabification as a post-lexical process comparable to post-lexical liaison in French and many other languages.

makes them similar to a group of unaccented morphemes which are generally referred to as *function words*, like the prepositions [az] 'from' and [be] 'to'. However, unlike clitics which are always cliticized leftward to a host, function words exhibit optionality in the direction of cliticization.[8,9]

The functional load of accentual contrast associated with clitics is extremely high at the level of PWORD. Two types of PWORD-level minimal pairs can be distinguished in terms of their morphosyntactic structure. First, segmentally identical MWORDs and MWORD-plus-clitic sequences contrast for the location of accent. The minimal pair in (2a) illustrates the contrast between a simple MWORD and a simple MWORD-plus-clitic, whereas (2b) gives the contrast between a derived MWORD and a simple MWORD-plus-clitic combination. Most clitics are homophonous with suffixes.

(2) a. vɒnet   vɒn = et
         (vɒ.nét)   (vɒ́.net)
         pickup   bathtub = POSS.2SG
                          'your bathtub'

     b. xub-i   xub = i
         (xu.bí)   (xú.bi)
         good-DER   good = COP.2SG
         'goodness'   'you are good'

Given that accentual contrasts of the sort just described arise from a difference in morphosyntactic structure, there are potentially countless numbers of minimal pairs of MWORD and MWORD-plus-clitic combinations in the language, especially in cases where suffixes are productive. An example of a highly productive suffix is the (colloquial) referential marker [-e] which is homophonous with the 3SG copula clitic [=e]. Thus, the referential nominal construction N + [-e], which is an MWORD, forms an accentual minimal pair with the predicate nominal construction N + [=e], which is an MWORD-plus-clitic, where the N slot can be filled by any noun/adjective of the language.

In the second type of minimal pairs, the two members are both MWORD-plus-clitic sequences. This situation arises in colloquial Persian only, since in such minimal pairs one member has undergone some morphophonological modification which only occurs in colloquial language. Specifically, some clitics have a particularly close relation to their MWORD host, as evidenced by the deletion of their initial vowel after an MWORD ending in vowel other than [e], thus resolving a potential vowel hiatus by the loss of a syllable. An example is the 1SG possessive clitic [=am] as in [labu] + [=am] → [labu = m] [labúm] 'my cooked beets'. A MWORD-final [e] is deleted in preference to the initial vowel of the clitic, as shown in (3a), whereby the accent appears on the vowel of the clitic (Ferguson 1957: 128). These forms contrast with segmentally identical MWORD-plus-clitic forms in which the MWORD ends in a consonant.

---

[8] To illustrate the flexibility in the direction of cliticization of prepositions, consider the preposition [az] in the sentential structure [mán az alí geréft = am] (I-from-Ali-caught.1SG) 'I caught it from Ali', which can be prosodically parsed either as (mán)(a.za.lí)(ge.réf.tam) or (má.naz)(a.lí)(ge.réf.tam). In highly careful speech, the preposition may form an independent PWORD: (mán)(az)(alí)(ge.réf.tam). Again, observe that the pattern of PWORD formation has no bearing on accent distribution. Thus the preposition is never accented, not even when it forms its own PWORD.

[9] For the sake of clarity, we do not use the term MWORD for unaccented syntactic elements like clitics and function words, reserving the term for major word classes as noted in Subsection 1.1.

(3) a.   xɒle = am → xɒlam   xɒl = am
           (xɒ.lám)   (xɒ́.lam)
           aunt = POSS.1SG   mole = POSS.1SG
           'my aunt'   'my mole'

     b.   bord-e = am → bordam   bord = am
           (bor.dám)   (bór.dam)
           took-PP = 1SG   took = 1SG
           'I have taken'   'I took'

The pair in (3b) illustrates the contrast between past and perfect verb forms. Since the perfect form consists of the past stem plus the participle suffix [-e] plus the personal clitic, all perfect forms minimally contrast with past forms which consist of past stem plus the personal clitic. Thus, these two verb systems are only distinct by the accent in colloquial Persian.[10] In formal/literary Persian the perfect forms are written as their underlying form and are pronounced as such. That is, the formal/literary equivalent of [bordám] is [bordéam].

Accentual minimal pairs are also highly frequent at the clause level, for instance due to the status of compounds as single MWORDs. To illustrate, the NP-VP clausal structure (4a) contrasts with the structure (4b), where [fɒrsi + zabɒn] is a compound.[11] Persian is a null-subject language.

(4) a.   fɒrsi   zabɒn = ast.
           (fɒr.sí)   (za.bɒ́.nast)
           Persian language = COP.3SG
           'Persian is a language.'
           [[N]$_{NP}$ [[N]$_{NP}$ V]$_{VP}$]$_S$

     b.   fɒrsi + zabɒn = ast.
           (fɒr.si)(za.bɒ́.nast)
           Persian + language = COP.3SG
           'S/he is a speaker of Persian.'
           [[[N + N]$_{NP}$ V]$_{VP}$]$_S$

Significantly, despite this extremely high functional load of accent location, Persian listeners have been found to be *stress deaf*, meaning that they perform poorly on reproducing the order of series of non-word stimuli that minimally differ in the position of phonetic prominence (e.g., [númi] and [numí]) (Rahmani et al. 2015). This is surprising if *stress deafness* is not expected to be found in languages with contrastive stress/accent, as suggested by Dupoux et al. (2001). However, as argued by Rahmani et al. (2015), if *stress deafness* in the short-term recall task relies on the listener's ability to store word prosody in their lexicon, the poor performance of Persian listeners can be explained by the assumption that Persian accent distinctions arise post-lexically, i.e. after words are inserted into phrase structure. Importantly, our data, as well as those of Rahmani et al. (2015), suggest that whereas these generalizations about accent assignment can be

---

[10] Note that this accentual contrast does not hold between 3sg forms of past/perfect since these forms have zero person marking (e.g., [bord-e] 'took-PP.3SG' vs. [bord] 'took.3SG').

[11] There have been some characterizations of sentential accent that require comment. Many researchers have described the final accent within a Persian sentence as more prominent than others, referring to it by such terms as "nuclear accent", "sentence stress", etc. The basis for these claims is unclear to us. There is no indication that the phonology of the language distinguishes, for example, between the two accents in (4a). In our view all accents at the phrase/clause level are equal.

transparently described in terms of morphosyntactic constituency, they are not obviously governed by the prosodic phrasing.[12,13]

In our experiment, we will use both types of minimal pairs that are formed at the level of PWORD, i.e. MWORD vs. MWORD-plus-clitic and MWORD-plus-clitic vs. MWORD-plus-clitic minimal pairs, exemplified above in (2) and (3) respectively. We will now move on to discuss two contexts that have been claimed to delete accent, the focus and presupposition constructions.

1.3 Focus

As in many languages, the variation in the focus of the sentence has intonational effects in Persian. Any MWORD in the sentence can be intonationally marked for focus, unless the MWORD is structurally unaccented within the sentential construction.[14] Variation in the focus of the sentence is illustrated for the subject-object-verb (SOV) structure in (5), which uses a question/answer paradigm. The size of the focus constituent is indicated by square brackets. The response in (5a) illustrates focus for the whole sentence. The responses in (5b) to (5d) illustrate focus on the subject, the object and the verb, respectively. In the response given in (5e), the subject and the object are both focused. The response in (5a) may be referred to as "neutral focus" in the sense that no constituent in the sentence is informationally more prominent than any other. All other responses, i.e. (5b) to (5e), can be seen as instances of "non-neutral focus".[15]

(5)   ɒlmɒn   junɒn = o   bord.
          Germany   Greece = OM   beat.3SG
          'Germany beat Greece.'

    a.   What happened?   [ɒlmɒn junɒn = o bord]$_{FOC}$
    b.   Who beat Greece?   [ɒlmɒn]$_{FOC}$ junɒn = o bord
    c.   Who did Germany beat?   ɒlmɒn [junɒn]$_{FOC}$ = o bord
    d.   What did Germany do to Greece? ɒlmɒn junɒn = o [bord]$_{FOC}$
    e.   Who beat whom?   [ɒlmɒn]$_{FOC}$ [junɒn]$_{FOC}$ = o bord

Figure 1 gives the F0 contour of the responses in (5a–5e), pronounced by a male speaker. As is obvious, post-focal constituents have a highly reduced F0 range. This has been widely observed in previous literature on Persian prosody (e.g., Mahjani 2003; Sadat-Tehrani 2007; Scarborough 2007), and been systematically studied by Taheri & Xu (2012), who described Persian as a *Post Focus Compression* (PFC) language. There is also evidence that

---

[12] Defining prosodic domains higher than PWORD does not appear to be as straightforward as in the case of PWORD. Our informal observations suggest that they may be characterized by different degrees of pause length and/or pre-boundary lengthening (cf. Mahjani 2003). Unfortunately, literature on Persian intonational structure does not provide a precise and systematic characterization of prosodic domains.

[13] The morphosyntactic nature of Persian accent is further supported by the observation that some grammatical words show variable accent position depending on the syntactic context. For instance, while the intensifier [xejli] 'very' is accented on the first syllable in positive sentences, e.g., [xéjli bad bud] 'It was very bad.', it takes accent on the final syllable in negative sentences, e.g., [xejlí bad nabud] 'It was not very bad.' (only the relevant accents are shown). Note that in both contexts the intensifier is parsed as a single PWORD.

[14] Persian is generally described as an SOV language, while allowing a great degree of rearrangements for pragmatic purposes. Crucially, post-verbal arguments and adverbials are structurally unaccented elements, and cannot be prosodically marked for focus (Karimi 2005; Sadat-Tehrani 2007). As an exceptional case, Sadat-Tehrani (2007: 138) observes that within "motion constructions", which involve predicates meaning 'go', 'come', etc., post-verbal elements are structurally accented and can also be prosodically focused.

[15] Example (5e) is a case of "double focus" (Scarborough 2007). While the "single focus" examples (5a) to (5d) focus the answer to wh-questions containing one wh-expression, the question to which (5e) is an answer contains two wh-expressions. We have no indication that this distinction affects the phonology of focused constituents in Persian.
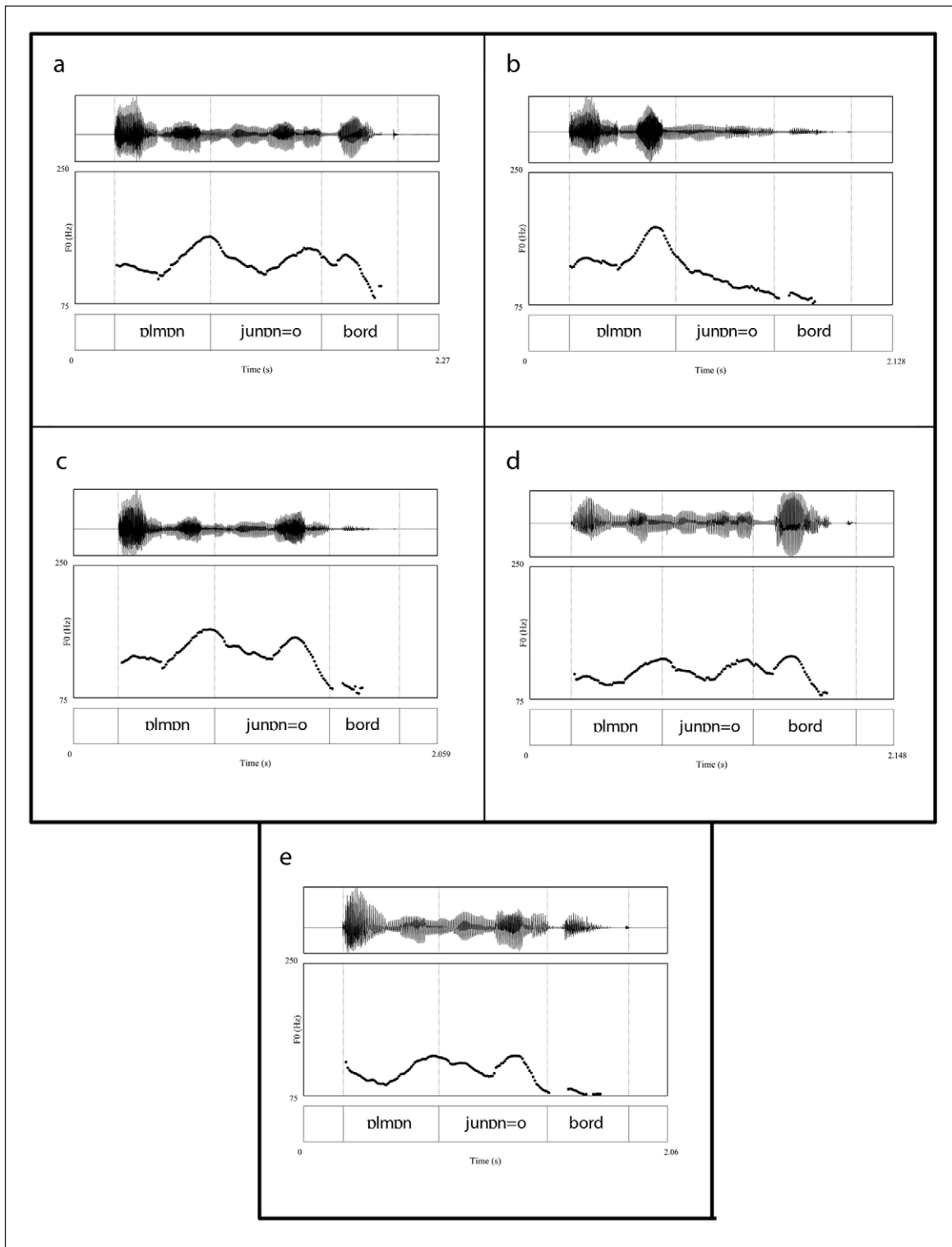
Figure 1: F0 contours, plotted below the associated waveforms, for the SOV structure [ɒlmɒn junɒno bord] 'Germany beat Greece', spoken by a male speaker in response to *What happened?* (panel (a)), *Who beat Greece?* (panel (b)), *Who did Germany beat?* (panel (c)), *What did Germany do to Greece?* (panel (d)), and *Who beat whom?* (panel (e)).

the presence of PFC is the most robust perceptual cue for focus (Taheri et al. 2014), a point that will be discussed more in Section 4.2.

The F0-lowering effect on the post-focal constituents is so significant that many researchers on the topic have claimed that the post-focal materials lose their accent to the extent that minimal accent pairs become homophonous (e.g., Eslami 2000; Vahidian-Kamyar 2001),

although these researchers reported no experimental data. The idea that constituents become deaccented after the focus, however, has been challenged by the experimental evidence in Abolhasanizadeh et al. (2012). We will turn to this issue in Subsection 1.5.

Another point evident from Figure 1 is that pre-focal MWORDs retain their accent, even if they are discourse-given, i.e. mentioned in the context question. Thus, responses in panel (a), where the whole sentence is focused, and in panel (d), where only the final MWORD is focused, have similar F0 shapes, although the MWORDs [ɒlmɒn] 'Germany' and [junɒn] 'Greece' are both given in the immediate context related to the latter. Moreover, when there are two focused MWORDs in the sentence (as in the response in panel (e)), the F0-compression starts after the second MWORD. The F0 shape of the response in panel (e) is thus comparable to that of in panel (c), where only the second MWORD is focused.

The minimal constituent for intonational marking of focus is the MWORD as noted before. Strikingly, a PWORD that is contained inside a MWORD cannot be intonationally marked for focus. For instance, the compound [fɒrsi + zabɒ́n] 'speaker of Persian', a single MWORD expressed as two PWORDs, is ambiguous between focus on the first member, on the second member, and on the whole construction, as shown in (6). However, these three focus domains may be distinguished by phonetic cues other than F0, as suggested by informal observation.[16]

(6) a.   [fɒrsi + zabɒn]$_{FOC}$
          (fɒr.si)(za.bɒ́n)
          Persian + language
          'speaker of Persian'

     b.   [fɒrsi]$_{FOC}$ + zabɒn
          (fɒr.si)(za.bɒ́n)

     c.   fɒrsi + [zabɒn]$_{FOC}$
          (fɒr.si)(za.bɒ́n)

The fact that Persian does not allow accentuation of the first constituent of a compound is not surprising given the status of accent as a morphosyntactic boundary marker, as shown for the noun + noun sequence [fɒrsi + zabɒn] (Persian + language) in (4). Thus, while [fɒrsí zabɒn ast] is an appropriate answer to the question *Is Iranian a language?*, it would be an inappropriate answer to *Is she a speaker of French?* That is, the phonological structure [fɒrsí zabɒn ast] can only be interpreted as *[Persian]$_{FOC}$ is a language* (cf. 4a), not as *She is a speaker of [Persian]$_{FOC}$*.

In addition to the possible phonetic effects of focus size, there have been two further distinctions that have been considered as sources of systematic phonetic variation. First, it has been noted that focused constituents have a relatively high F0 (e.g., Sadat-Tehrani 2007; Taheri & Xu 2012). In particular, Scarborough (2007) suggested that the accent for focus is part of the phonological representation, resulting from a H^ (superhigh) tone, as opposed to the neutral H tone. The evidence that we present in the current study suggests

---

[16] A similar case has been reported for French Noun + Adjective sequences by Hamlaoui et al. (2012), as noted by an anonymous reviewer. Hamlaoui et al. (2012) argued that prosodic phrasing might be a factor in the French data. The same may well be true for Persian. For instance in (6b) the focus constituent, i.e. the first component of the compound, may form an independent prosodic phrase manifested with a relatively high degree of strengthening and a strong pause between the components of the compound. This process has no effect on the distribution of accent, given the morphosyntactic requirement that the right edge of the MWORD, i.e. the compound as a whole, be accented.

that there is no phonological difference between the accent for focus and the neutral accent, and that F0-raising for focus can be attributed to paralinguistic speaker strategies. Second, focus type has been claimed to have an effect. The question/answer pairs in (5) illustrate *presentational focus*, a term that has been used for information provided by the speaker in reply to the hearer's request, either overt or implied. Previous studies about Persian focus have also been concerned with other types of focus, in particular *contrastive focus*, which relates the focus constituent to a restricted set of alternatives (cf. Gussenhoven 2008). In her syntactic approach to Persian focus, Karimi (2005: 132) claimed that contrastive focus and presentational focus show different prosodic realizations, although she is not clear as to exactly what the difference is.[17] We are not aware of any experimental study that systematically compares intonational effects of presentational focus and contrastive focus in Persian, but the existing data do not indicate any obvious intonational differences between these focus types (Vahidian-Kamyar 2001; Mahjani 2003; Scarborough 2007).

1.4 Factive verbs and presupposed dependent clauses

Persian has certain constructions that are semantically related to focus and which have similar intonational effects. Scarborough (2007) has described them as "focus-like" constructions which are triggered by specific syntactic contexts or are determined by the lexical semantics of particular items. She notes that the prosody of these constructions resembles the prosody of focused constituents, as they are "marked by a high pitch accent and followed by deaccenting" (Scarborough 2007: 28). Examples are wh-questions, negation constructions and clausal complements of certain predicates such as [dunestan] (realized as [dɒnestan] in formal styles) 'to know'. Here we limit our discussion to the latter.

Scarborough observed two intonation patterns in the pronunciation of sentences containing complement clauses, exemplified in (7), where complement clauses are in brackets. Persian clausal complements, which appear in post-verbal position, may be introduced by the morpheme [ke].[18]

(7)   a.   mɒdarbozórg   mídune   [(ke)   ɒnɒnɒse   reside].
               Grandmother   DUR.know.3SG   that   pineapple.ART   ripen.PP.3SG
               'Grandmother knows that the pineapple has ripened.'

        b.   mɒdarbozórg   migé   [(ke)   ɒnɒnɒsé   residé].
               grandmother   DUR.say.3SG   that   pineapple.ART   ripen.PP.3SG
               'Grandmother says that the pineapple has ripened.'

The author argues that while (7b) follows the neutral intonation of Persian declaratives in the sense that all MWORDs are accented, in (7a) the verb form [midune] 'she knows' is pronounced as if it was focused, since the MWORDs in the complement clause are deaccented, even though speakers produced this intonation in a contextless reading. Scarborough suggested that this "focus-like" pronunciation might be due to an "obligatory semantic focus" as triggered by the lexical semantics of a particular verb class, of which the verb [dunestan] 'to know' is an example. She provided no further explanation of this

---

[17] Karimi (2005) describes that Persian contrastive focus and presentational focus (or "information focus", in her terminology) are prosodically distinguished in that the former, unlike the latter, carries a "heavy stress". She does not however clarify the term "heavy stress" (see also footnote 11).

[18] The non-final position of accent in the verb form [mídune], which is a syntactically complex element (cf. Kahnemuyipour 2003), is irrelevant to our discussion.

observation. We also refer the reader to Sadat-Tehrani (2007) on these effects for verbs meaning 'to forget', the only other reference to this phenomenon in the literature as far as we are aware.

The semantic contrast between the main verbs in (7a) and (7b) may remind us of a traditional distinction between two classes of clause-taking verbs, factives (e.g., *know*, *realize*, *regret*) and non-factives (e.g., *say*, *believe*, *think*). Factive verbs presuppose the truth of their complement clause while non-factive verbs do not (Kiparsky & Kiparsky 1970). For example, in *she knows that the cat is in the garden*, the speaker presupposes that the cat is in the garden, while in *she says that the cat is in the garden,* there is no such requirement (Crystal 2008: 184).

We will refer to deaccented clausal complements of factive verbs as *presupposed* clauses. In (7a), where the embedded clause is deaccented, the ripeness of the pineapple is presupposed, while no such presupposition is present in the accented clausal complement of (7b). While we do not deny that in an out-of-the-blue context clausal complements of factive verbs are more likely to be deaccented as suggested by Scarborough's data, they can be accented under specific discourse conditions, specifically when they contain non-presupposed information.[19] Example (8) may be spoken in reply to *What is the latest news from the World Cup?* In this context, the complement clause does not convey presupposed information in the sense discussed above.

(8)   mídunam   [ɒlmɒ́n   junɒ́n = o   bórd].
         DUR.know.1SG   Germany   Greece = OM   beat.3SG
         'I know (that) Germany beat Greece.'

Crucially, the main verb [midunam] does not merely express factivity, but is used to indicate that the information in the complement clause is only a partial answer to the question, a meaning which has been categorized as *non-exhaustive focus* by Elordieta & Irurtzun (2010). We may paraphrase (8) as follows:

> "I am not in a position to give a full answer to your question. But here is a relevant fact which I do know: Germany beat Greece."

A syntactic correlate of the distinction between the non-exhaustive complement clause (8) and the presupposed one in (7a) is that only the non-exhaustive one allows preposing, as illustrated in (9).[20] When this happens, the (normally factive) verb loses its accent.

(9)   [ɒlmɒ́n junɒ́n = o bórd] midunam.
         'Germany beat Greece, I know.'

These observations are of practical relevance to our experiment. The fact that clausal complements of factive verbs can be accented as well as deaccented depending on the meaning makes them suitable as stimuli in our experiment.

1.5 *The problem*

Abolhasanizadeh et al. (2012), using both production and perception data, have shown that while the F0 range of their target minimal pairs is considerably reduced after the focus, the tonal structure remains intact, as reflected in above-chance identification of

---

[19] It has widely been observed that English factive predicates may take non-presupposed (asserted) complements in certain contexts (e.g., Hooper & Thompson 1973; Simons 2007).
[20] This phenomenon is comparable to the English "Slifting (Sentence lifting) constructions" (Ross 1973).

the post-focal items by the listeners. Thus, the accentual difference between the Mᴡᴏʀᴅ [tɒbéʃ] and the Mᴡᴏʀᴅ-plus-clitic [tɒ́b = eʃ] was preserved in the post-focal context in (10).

(10)   a.   [un]_ꜰᴏᴄ   tɒbéʃ = e.
                 that   Tabesh = ᴄᴏᴘ.3 sɢ
                 'That is Tabesh.'
        b.   [un]_ꜰᴏᴄ   tɒ́b = eʃ = e.
                 that   loll swing = ᴘᴏss.3sɢ = ᴄᴏᴘ.3sɢ
                 'That is his/her loll swing.'

Our interest in the present study is to reconsider this issue. We believe that the validity of the data used by Abolhasanizadeh et al. (2012) is questionable. In fact, the status of *focus* is unclear in their production design, which in turn yielded the stimuli for their perception experiment. They used a reading task in which the intended focused elements were printed in bold letters. Since there was no question/answer paradigm, it is unclear how each speaker interpreted information status of the bold elements in the text. It is possible that at least some speakers pronounced the bold items merely in a more emphatic way, i.e. with more acoustic energy. This is important because emphasis, unlike focus, does not create post-focal deaccenting in Persian (we will provide more information on the distinction between focus and emphasis in Subsection 4.3). This consideration motivated us to replicate Abolhasanizadeh's study with a more realistic elicitation task and a larger corpus. The specific objective of our study was thus to determine whether Persian words undergo deaccentuation. We investigated this question through the use of two experiments, a production experiment and a perception experiment.

In order to place the investigation in a wider perspective, we decided to include different putative deaccenting contexts in our corpus. As target items, we used two minimal pairs with different morphological structures. The target minimal pairs were each examined in the context of the two types of information structural configurations, namely contrastive focus and factive verb construction described above in Subsections 1.3 and 1.4.

2 Experiment I

We conducted a production experiment to gather phonetic data on the realization of accentual contrasts in Persian. The strategy of the experiment was to have speakers pronounce a set of pre-designed sentences in different information structure contexts. In the current study, we are only concerned with measures of fundamental frequency (F0).

2.1 Materials

We built up a corpus of sentences featuring two minimal pairs as target items, given in Table 1. The first pair is a contrast between a simple noun and a noun-plus-clitic, while the second pair illustrates a contrast between two verb-plus-clitic sequences.[21] Target items that are initially accented are referred to as initial accent condition, while those with final accent are referred to as final accent condition.

We constructed four carrier sentence structures, each corresponding to one of the possible combinations of the two minimal pairs ([tɒ́beʃ/tɒbéʃ], [dídan/didán]) and the two information structural configurations (contrastive focus, factive verb construction). The carrier sentences were designed in such a way that they provide the minimal context required to cover all experimental conditions in which they were used. Table 2 gives the

---

[21] The first minimal pair, i.e. [tɒ́beʃ/tɒbéʃ], is the same as that used in Abolhasanizadeh et al. (2012).

Table 1: The minimal pairs as target items used in the experiment.

	First member (initial accent)	**Second member (final accent)**
1ˢᵗ pair	tŏb=eʃ loll swing=POSS.3SG 'his loll swing'	tɒbéʃ proper name 'Tabesh'
2ⁿᵈ pair	díd=an saw=3PL 'they saw'	did-án saw-PP.3PL 'they have seen'

Table 2: The structure of the corpus used in the experiment.

Information structural configuration	**Target pair**	**Carrier sentence (X denotes the target word)**
Contrastive focus (neutral/focal/post-focal)	[tŏbeʃ/tɒbéʃ] 'his swing/Tabesh'	*un X=e* 'That is X.'
Contrastive focus (neutral/focal/post-focal)	[dídan/didán] 'they saw/they have seen'	*unɒ X=eʃ* 'They X it.'
Factive verb construction (non-presupposed/presupposed)	[tŏbeʃ/tɒbéʃ] 'his swing/Tabesh'	*midunam X=e* 'I know (that) this is X.'
Factive verb construction (non-presupposed/presupposed)	[dídan/didán] 'they saw/they have seen'	*midunam X=eʃ* 'I know (that) they X it.'

structure of the corpus. There were three experimental conditions for focus, (1) neutral (out-of-the-blue pronunciation), (2) focal (the target item is contrastively focused), and (3) post-focal (the MWORD preceding the target item is contrastively focused). For factive verb construction, there were two experimental conditions, (1) non-presupposed (the target item is part of a non-presupposed complement clause) and (2) presupposed (the target item is part of a presupposed complement clause).

Within the carrier sentences, the target items had a cliticized form, i.e. $X=e$ for the structures involving the minimal pair [tŏbeʃ/tɒbéʃ], and $X=eʃ$ for the structures with the minimal pair [dídan/didán], which makes them all part of trisyllabic PWORDs that contrast in having the accent on the antepenultimate syllable (for initial accent target items) or on the penultimate syllable (for final accent target items). The motivation to use a PWORD-final unaccented syllable in all cases was to avoid local phrase-final effects on the realization of the target syllables.

To elicit the desired experimental conditions, carrier sentences were preceded by a context sentence, with which they formed mini-dialogues. This resulted in 12 mini-dialogues for focus and eight mini-dialogues for factive construction (see Table 3 for examples; all mini-dialogues are given in Appendix 1). Example mini-dialogues, which were different from the test materials, were provided to the speakers to practice each condition before the actual recording. There were three example dialogues for focus and two example dialogues for factive construction, which were recorded by two male speakers different from the test speakers.

Table 3: English translations of the mini-dialogues that were used to elicit experimental conditions for the target item [tɒbéʃ].

Contrastive focus	Neutral	A: What happened? B: That is Tabesh.
	Focal	A: Is that Ahmadi? B: That is Tabesh.
	Post-focal	A: Is this Tabesh? B: That is Tabesh.
Factive verb construction	Non-presupposed	A: Have you heard about the new teacher? B: I know he is called Tabesh.
	Presupposed	A: He says his name is Tabesh but I don't think so. B: I know he is called Tabesh.

2.2 Speakers

Eight speakers took part in the experiment, four male and four female, aged from 27 to 37. They were native speakers of Standard Persian, all with university education. Informed consent was obtained from each participant.

2.3 Procedure and recording

The dialogues were presented to the speakers in two experimental blocks. One block consisted of materials related to contrastive focus, while the materials related to factive verb construction were put in the other block. The order of blocks and of test dialogues within each block was randomized per speaker in order to control for presentation order. Each block was presented to the speakers in the form of a booklet with one dialogue per page in Persian orthography. Mword-plus-possessive clitics can be written either as one orthographic word (the traditional standard) or as separate orthographic words. We wrote [tɒ́beʃ] as two orthographic words to distinguish it from [tɒbéʃ]. Since there is no standard way of distinguishing between [dídan] and [didán] in Persian orthography, these items were written in their formal/literary form, which corresponds to [didand] and [dideand], respectively. The person/number marker [=an] (3PL) is realized as [=and] in formal and literary Persian.

Before the recording of each block, speakers were given an opportunity to try out the example dialogues by listening to the sound files played from a laptop and pronouncing the target sentence, for as long as they wished. In general, the speakers found the task clear and easy and quickly moved on from the trial.[22] During the recording, the experimenter pronounced the context sentences and the speakers read aloud the response sentence from the booklet, without being able to see the experimenter's face.[23] The procedure was performed twice to obtain two recordings from each condition per participant.

---

[22] A reviewer expressed concern that the use of example recordings might have amounted to an imitation task. Our purpose for using example dialogues was to ensure that participants understood the task. We followed a common procedure to give an example of a stimulus and a possible response and gave participants an opportunity to see if they had any questions. There was therefore no specific training of subjects as to how to pronounce the materials. We preferred this method over the alternative of discarding erroneous readings, as this would have required a criterion that might have pre-judged the results of the experiment. With the exception of a few evident disfluencies, all productions by our subjects were processed. Our procedure was uncomplicated and no skills were required that go beyond what the average naïve speaker of the language can do without training.

[23] Given that the context sentences were spoken during the experimental session by the experimenter, one reviewer asked how variability between different experimental sessions was controlled for. The contexts sentences were either interrogatives or declaratives and were consistently pronounced by the experimenter

Unlike previous studies which used bold letters for focused items (e.g., Abolhasanizadeh et al. 2012), we did not highlight focused items in any way, assuming that the appropriate context questions should suffice to elicit the intended focus structure. Highlighting the intended focus elements typographically might moreover have caused subjects to pronounce these items in particularly emphatic ways. We return to this point in Section 4.

The speakers were recorded individually in the studio of the Linguistics Department of the University of Tehran using a Shure SM58 vocal cardioids microphone (44.1 kHz, mono channel, 16-bit). For analysis, one recording was randomly selected from the utterances of each sentence by each speaker which resulted in 20 response utterances per speaker overall.

2.4 Results

We report general observations on the effects of information structure on the realization of accentual contrasts based on visual inspection of time-normalized averaged F0 curves pooled over all eight speakers, followed by statistical evaluation with analysis of variance for repeated measures (rmANOVA). A Praat script was used to carry out the measurements and to correct pitch errors resulting from creaky voice or octave jumps (Boersma & Weenink 2014).

F0 measurements for initial accent and final accent target items will be reported separately for each of the four parts of the corpus given in Table 2. Figures 2 and 3 show averaged contours on normalized time scales for the minimal pairs [tǿbeʃ/tɒbéʃ] and [dídan/didán] in the three focus conditions, while Figures 4 and 5 show the two factive conditions. Time-normalized F0 measurements were collected at 12 equidistant points per syllable.

We turn to the results for focus first. Panels (a) in Figures 2 and 3 present the contours in the neutral condition for [tǿbeʃ/tɒbéʃ] and [dídan/didán], respectively, which show the effect of the position of the accent. The initial syllable is 3.83 ST (SD = 1.95) ([tǿbeʃ/tɒbéʃ]) and 3.09 ST (SD = 1.68) ([dídan/didán]) higher when accented than its unaccented counterpart. For the final syllable, the difference is 2.41 ST (SD = 1.24) and 2.80 ST (SD = 2.31) for [tǿbeʃ/tɒbéʃ] and [dídan/didán], respectively. Panels (b) present the contours in the focal condition, which show a comparable pattern. The initial syllable here is 4.56 ST (SD = 1.66) ([tǿbeʃ/tɒbéʃ]) and 4.21 ST (SD = 1.36) ([dídan/didán]) higher, and the final syllable is 3.37 ST (SD = 1.71) ([tǿbeʃ/tɒbéʃ]) and 2.74 ST (SD = 1.04) ([dídan/didán]) higher when accented than when unaccented. Panels (c) present the contours for the post-focal condition. A different pattern is observed in the low F0 plateau for both minimal pairs, which would appear to show that the post-focal items are deaccented.[24]

Turning to the effect of factive verb construction, we observe a similar pattern of deaccentuation for both minimal pairs. While a considerable F0 difference exists between the members of each minimal pair in the non-presupposed condition (panels (a) in Figures 4 and 5), there is a low plateau on the target items in the presupposed condition (panels (b) in Figures 4 and 5). In the non-presupposed condition, the F0 difference between the initial accented and the initial unaccented syllables is 4.51 ST (SD = 1.80) ([tǿbeʃ/tɒbéʃ]) and 3.60 ST (SD = 1.79) ([dídan/didán]), whereas for the final syllables the difference is 2.67 ST (SD = 1.29) ([tǿbeʃ/tɒbéʃ]) and 3.16 ST (SD = 0.93)

---

with unmarked and grammatically correct intonation. Phonetic variability between different sessions was considered irrelevant for the purposes of this study.

[24] Mean F0 differences between the target syllables in the post-focal condition are as follows. For the initial syllables, the difference is 0.09 ST (SD = 1.01) ([tǿbeʃ/tɒbéʃ]) and 0.55 ST (SD = 2.31) ([dídan/didán]), while the difference between the final syllables is 0.02 ST (SD = 0.74) ([tǿbeʃ/tɒbéʃ]) and 0.14 ST (SD = 0.64) ([dídan/didán]).
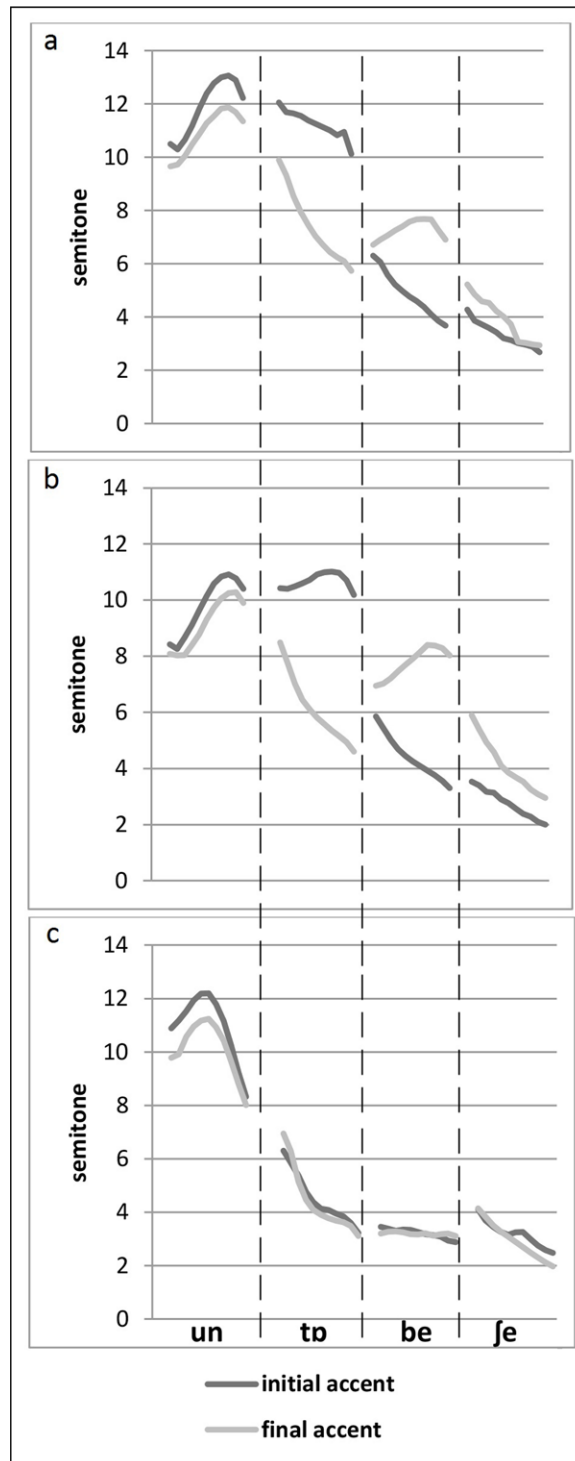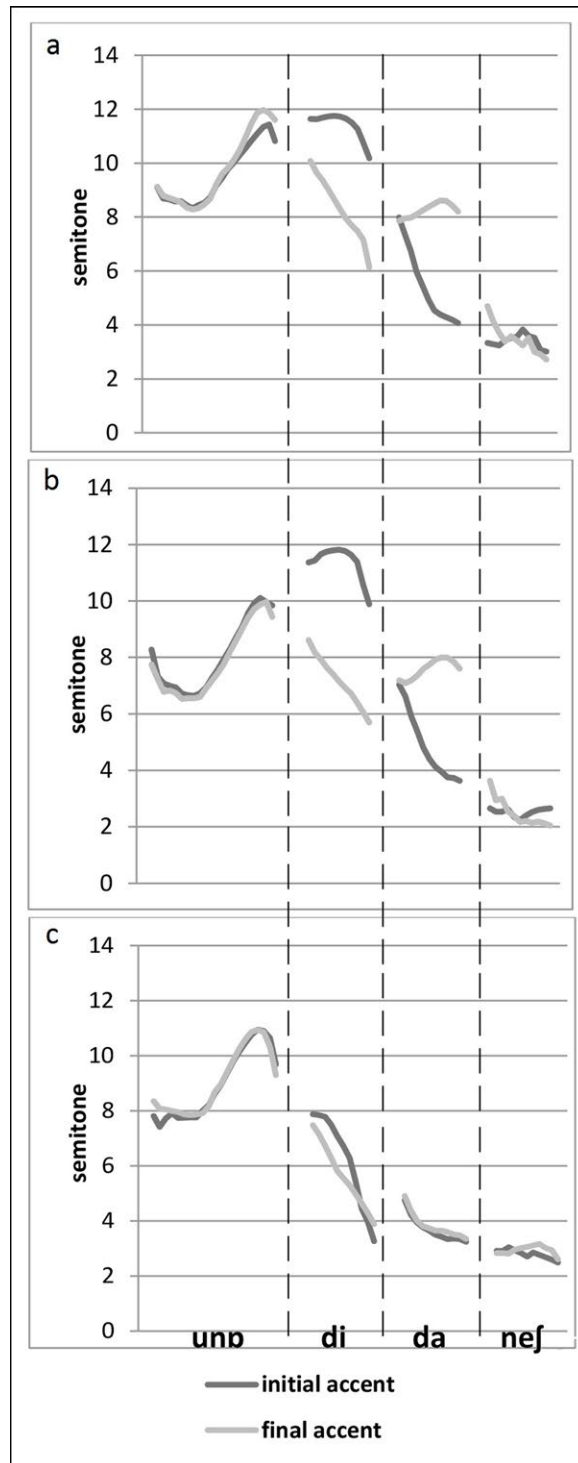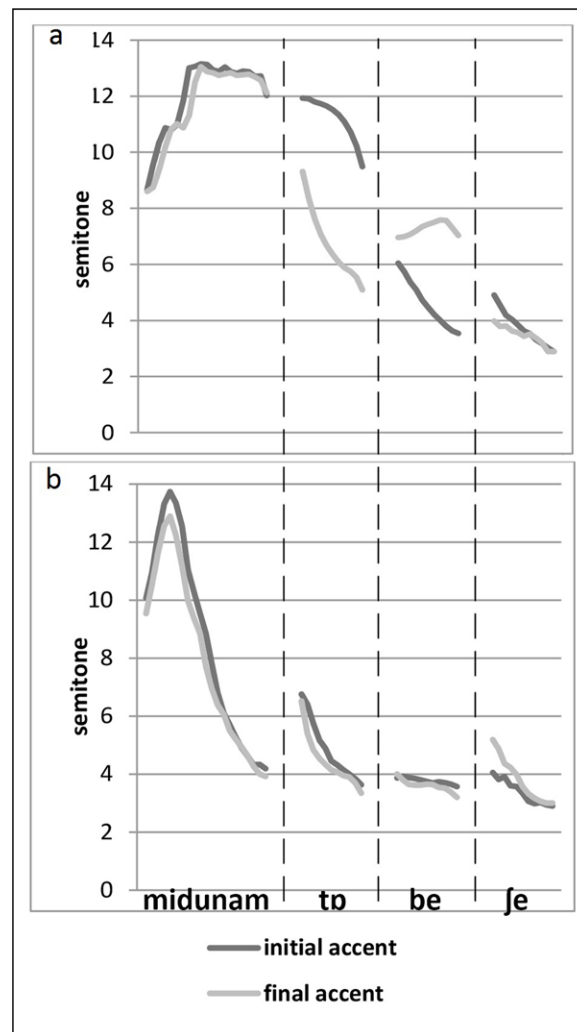
Figure 2: Mean F0 contours for [un tɒbeʃe] on a normalized time scale for initial accent and final accent items separately, with target items in neutral pronunciation (panel (a)), in focal pronunciation (panel (b)), and post-focal pronunciation (panel (c)), pooled over 8 speakers. The values are given in semitones (reference: 100 Hz).

([dídan/didán]).[25] Thus, the presupposed condition appears to have the same deaccenting effect as the post-focal condition of contrastive focus.

---

[25] Mean F0 differences between the target syllables in the presupposed condition are as follows. For the initial syllables, the difference is 0.42 ST (SD = 0.94) ([tɒ́beʃ/tɒbéʃ]) and 0.08 ST (SD = 0.62) ([dídan/didán]), while for the final syllables the difference is 0.16 ST (SD = 0.41) ([tɒ́beʃ/tɒbéʃ]) and 0.36 ST (SD = 0.87) ([dídan/didán]).

Figure 3: Mean F0 contours for [ʊnɒ didaneʃ] on a normalized time scale for initial accent and final accent items separately, with target items in neutral pronunciation (panel (a)), in focal pronunciation (panel (b)), and post-focal pronunciation (panel (c)), pooled over 8 speakers. The values are given in semitones (reference: 100 Hz).

In order to find statistical evidence for these results, we conducted eight repeated measures ANOVAs. Each of these included data from either the first or the second syllable of the members of a minimal pair from either the focus or the factive part of the experiment, i.e. 2 minimal pairs × 2 syllable positions × 2 information structure

Figure 4: Mean F0 contours for [midunam tɒbeʃe] on a normalized time scale for initial accent and final accent items separately, with target items in non-presupposed pronunciation of the embedded clause (panel (a)), and presupposed pronunciation of the embedded clause (panel (b)), pooled over 8 speakers. The values are given in semitones (reference: 100 Hz).

constructions. The independent variables were information structure (neutral, focal, post-focal for contrastive focus; non-presupposed, presupposed for factive verb construction) and accent position (initial accent, final accent). The dependent variable was the mean over the six F0 values extracted from the second half of each syllable. This limitation was applied so as to minimize carryover effects of preceding syllables. The analyses are reported in Tables 4 to 7. Figures 6 to 9 show the corresponding mean values.

In all analyses, the main effects of accent position and information structure were found to be significant (with an alpha level set to 0.05) with large effect sizes. Overall, first, accented syllables of target items have significantly higher F0 than unaccented ones, and, second, F0 is significantly lower in post-focal and presupposed target items than in the other conditions. Moreover, all ANOVAs revealed significant interactions of information structure and accent position with relatively large effect sizes, due to the fact that the F0 difference between accented and unaccented syllables is not significant in post-focal and presupposed conditions. These results confirm the expectation expressed above.
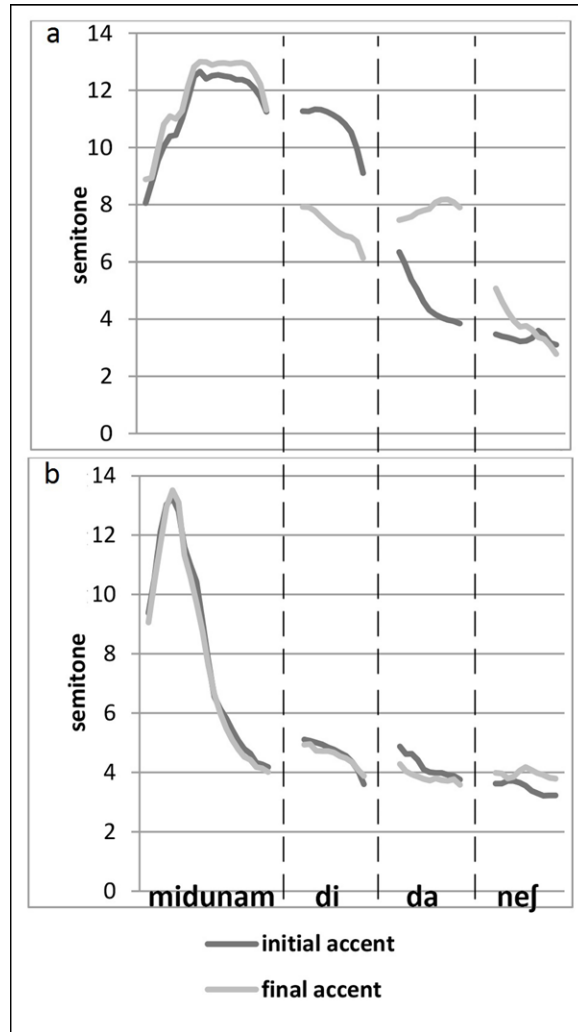
Figure 5: Mean F0 contours for [midunam didaneʃ] on a normalized time scale for initial accent and final accent items separately, with target items in non-presupposed pronunciation of the embedded clause (panel (a)), and presupposed pronunciation of the embedded clause (panel (b)), pooled over 8 speakers. The values are given in semitones (reference: 100 Hz).

3 Experiment II

The results of Experiment I suggested that the tonal distinctions between the members of the accentual minimal pairs are lost in the post-focal condition of contrastive focus and in the presupposed condition of factive verb construction. To verify this observation, a perception experiment was designed to test the hypothesis that the members of our minimal pairs are homophonous in these post-focal and presupposed conditions.

3.1 Materials

All response sentences analyzed in Experiment I were included in the corpus of Experiment II. The utterances were divided into two blocks, one of which contained sentences related to contrastive focus, and the other contained sentences related to factive verb construction. This resulted in 160 stimuli. The design is given in Table 8.

Table 4: ANOVA summary of the effects of contrastive focus and accent position on F0 measures of the 2nd half of the syllable [tɒ] and the 2nd half of the syllable [be] of the sequence [tɒbeʃe]. IS: information structure; AP: accent position.* indicates a significant effect at the 5% level, Huynh-Feldt corrected where appropriate.

	Effects	F	p	Partial η²	Post-hoc comparison, Sidak's p
[tɒ]	IS	$F_{2,14}$ = 118.84*	<.001	.944	
	Neutral vs. Focal				.148
	Neutral vs. Post-focal				<.001*
	Focal vs. Post-focal				<.001*
	AP	$F_{1,7}$ = 45.94*	<.001	.868	
	IS × AP	$F_{2,14}$ = 32.02*	<.001	.821	
[be]	IS	$F_{2,14}$ = 24.04*	<.001	.775	
	Neutral vs. Focal				.911
	Neutral vs. Post-focal				.004*
	Focal vs. Post-focal				.004*
	AP	$F_{1,7}$ = 134.96*	<.001	.951	
	IS × AP	$F_{2,14}$ = 29.03*	<.001	.806	

Table 5: ANOVA summary of the effects of contrastive focus and accent position on F0 measures of the 2nd half of the syllable [di] and the 2nd half of the syllable [da] of the sequence [didaneʃ]. IS: information structure; AP: accent position.* indicates a significant effect at the 5% level, Huynh-Feldt corrected where appropriate.

	Effects	F	p	Partial η²	Post-hoc comparison, Sidak's p
[di]	IS	$F_{2,14}$ = 100.40*	<.001	.935	
	Neutral vs. Focal				.581
	Neutral vs. Post-focal				<.001*
	Focal vs. Post-focal				<.001*
	AP	$F_{1,7}$ = 47.85*	<.001	.872	
	IS × AP	$F_{2,14}$ = 18.06*	<.001	.721	
[da]	IS	$F_{2,14}$ = 13.98*	<.001	.666	
	Neutral vs. Focal				.656
	Neutral vs. Post-focal				.006*
	Focal vs. post-Focal				.028*
	AP	$F_{1,7}$ = 61.95*	<.001	.898	
	IS × AP	$F_{2,14}$ = 18.76*	<.001	.728	

3.2 Listeners

A total of 21 listeners, different from the speakers in Experiment I, were recruited, 11 male and 10 female, aged from 17 to 45. They were native speakers of standard Persian,

Table 6: ANOVA summary of the effects of factive verb construction and accent position on F0 measures of the 2nd half of the syllable [tɒ] and the 2nd half of the syllable [be] of the sequence [tɒbeʃe]. IS: information structure; AP: accent position.* indicates a significant effect at the 5% level, Huynh-Feldt corrected where appropriate.

	Effects	F	p	Partial η²
[tɒ]	IS	$F_{1,7}$ = 86.65*	<.001	.925
	AP	$F_{1,7}$ = 57.90*	<.001	.892
	IS × AP	$F_{1,7}$ = 24.64*	.002	.779
[be]	IS	$F_{1,7}$ = 32.36*	.001	.822
	AP	$F_{1,7}$ = 62.60*	<.001	.899
	IS × AP	$F_{1,7}$ = 52.51*	<.001	.882

Table 7: ANOVA summary of the effects of factive verb construction and accent position on F0 measures of the 2nd half of the syllable [di] and the 2nd half of the syllable [da] of the sequence [didaneʃ]. IS: information structure; AP: accent position.* indicates a significant effect at the 5% level, Huynh-Feldt corrected where appropriate.

	Effects	F	p	Partial η²
[di]	IS	$F_{1,7}$ = 129.60*	<.001	.949
	AP	$F_{1,7}$ = 25.36*	.002	.784
	IS × AP	$F_{1,7}$ = 38.88*	<.001	.847
[da]	IS	$F_{1,7}$ = 21.50*	.002	.754
	AP	$F_{1,7}$ = 44.81*	<.001	.865
	IS × AP	$F_{1,7}$ = 87.60*	<.001	.926

all of whom had obtained a university degree. None of them reported hearing problems and informed consent was obtained from each of them.

3.3 Procedure

The experimental task was presented with a Praat Multiple Forced Choice interface on a laptop (Boersma & Weenink 2014). The stimuli were played through headphones. Listeners were told that they were going to hear a series of sentences, each containing one of the four items [tɒbeʃ] 'his swing', [tɒbéʃ] 'proper name', [dídan] 'they saw' and [didán] 'they have seen', and that they should select which of these items each sentence contained. As in Abolhasanizadeh et al. (2012), these items appeared on the screen in Persian orthography (in the same forms as described in Subsection 2.3) in four clickable boxes.

Within each block, the order of the stimuli was randomized per listener, while the order of blocks was counterbalanced across the listeners. Before the test, listeners were given eight trial stimuli to familiarize themselves with the task. These were selected from the neutral and non-presupposed conditions, in which the target items were accented, so as to facilitate their recognition. All participants indicated that they thought the task was clear to them. They were told that during the experiment they could listen to each stimulus as often as they wished, but that once they had made their choice, it could not be changed.

Participants were tested in different places in Tehran: ten in their home, four in libraries and eight in the studio of the Linguistics Department of the University of Tehran.
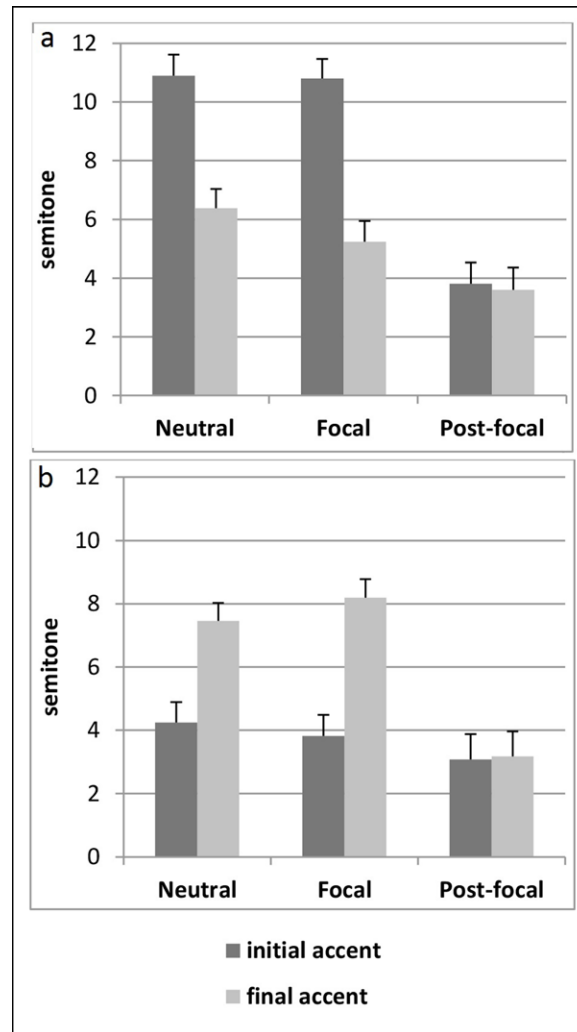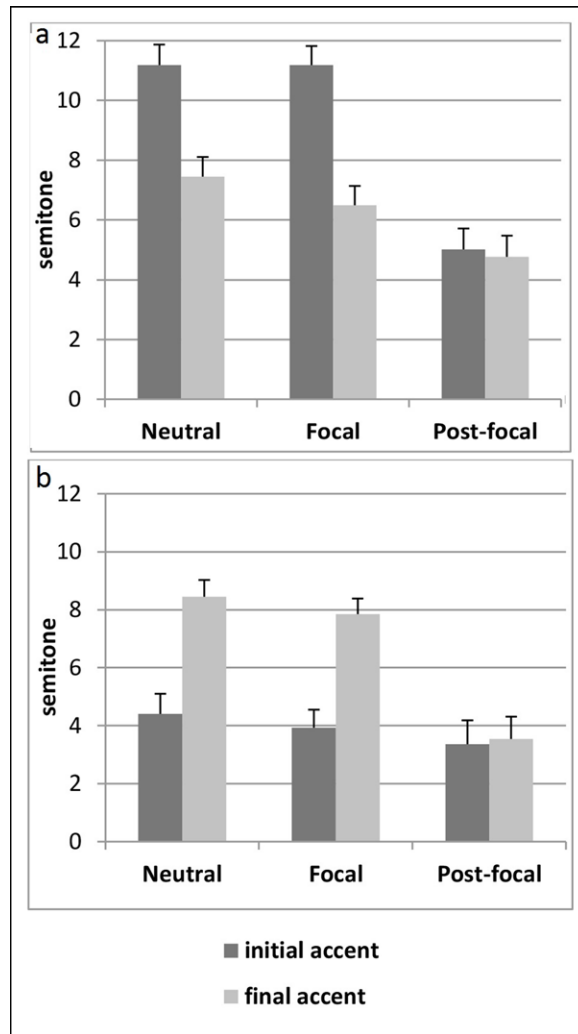
Figure 6: Mean F0 values of the 2nd half of the syllable [tɒ] (panel (a)) and 2nd half of the syllable [be] (panel(b)) for initial accent and final accent items within the carrier structure [tɒbeʃe] in neutral, focal and post-focal conditions of contrastive focus, pooled over 8 speakers. The values are given in semitones (reference: 100 Hz). Error bars represent standard error of the mean.

3.4 Results

In our analysis, a response was considered correct if the listener identified the item that the speaker in the production experiment was supposed to produce. To obtain a quantitative measure of correct identification, we used *d*-prime (*d′*), a sensitivity index used in Signal Detection Theory (Macmillan & Creelman 1991). This is given by the equation in (11):

(11) $d' = z(H) - z(F)$

H is the proportion of correct responses to one stimulus type (*hit* rate), *F* is the proportion of incorrect responses to the other stimulus type (*false alarm* rate), and $z()$ gives the *z*-score of these variables. The measure *d′* is suitable for our purposes, because it eliminates any biases in the response rates that may have arisen due to the decision rules a subject uses. For our identification data, we calculated *hits* and *false alarms* for individual minimal pairs in each experimental block. *d′* values were obtained for each listener. In general, higher *d′* scores indicate better performance. In our analysis, we took *d′* = 1.35 to be the baseline performance. This *d′* value corresponds with correct performance on 75% of the
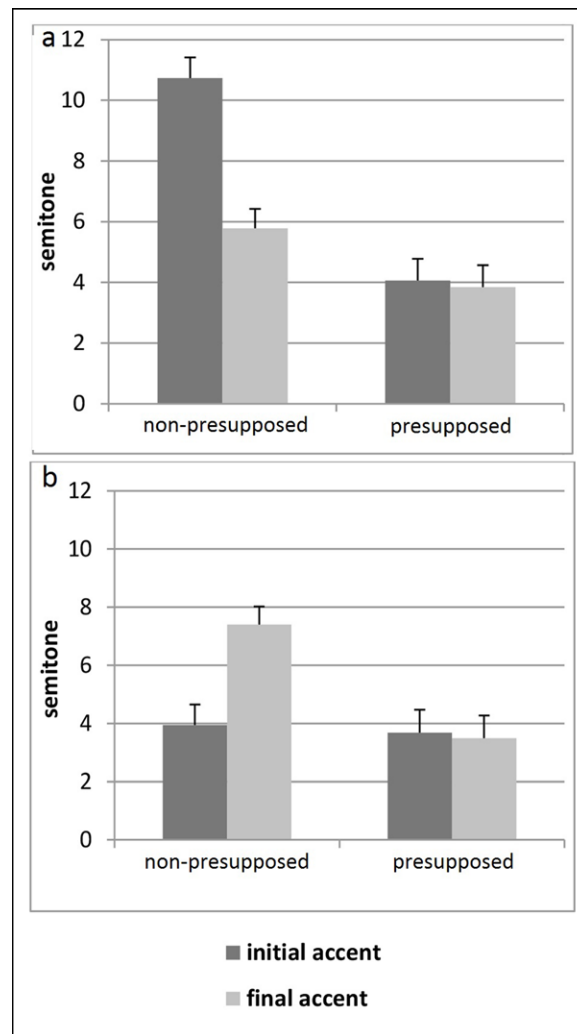
Figure 7: Mean F0 values of the 2nd half of the syllable [di] (panel (a)) and 2nd half of the syllable [da] (panel(b)) for initial accent and final accent items within the carrier structure [didaneʃ] in neutral, focal and post-focal conditions of contrastive focus, pooled over 8 speakers. The values are given in semitones (reference: 100 Hz). Error bars represent standard error of the mean.

trials (Macmillan & Creelman 1991), which is often seen as the correct rate for a just-noticeable-difference (JND) (Durrant & Lovrinic 1977).

The maximum possible d' score was 3.06 and the range of scores that the listeners achieved extended from –1.15 to 3.06. In line with common practice (Macmillan & Kaplan 1985), where d' would otherwise be undefined (a hit or false alarm rate of zero or 1), rates of 0 were replaced with $[0.5/n]$, and rates of 1 were replaced with $[(n – 0.5)/n]$, where n is the maximum number of hits or false alarms. For our data, in which n = 8, these values were 0.062 and 0.937, respectively.

We report averaged d' values with corresponding standard errors pooled over 21 listeners in Figures 10 and 11 for contrastive focus and factive verb construction, respectively. There were no missing data in our analysis. Nor was it ever the case that an item was picked from the wrong minimal pair.

d' values for each block were subjected to a separate repeated-measures ANOVA with minimal pair ([tʊ́beʃ/tʊbéʃ], [dídan/didán]) and information structure (neutral, focal, post-focal) as factors for contrastive focus, and minimal pair ([tʊ́beʃ/tʊbéʃ], [dídan/didán])

Figure 8: Mean F0 values of the 2nd half of the syllable [tɒ] (panel (a)) and 2nd half of the syllable [be] (panel (b)) for initial accent and final accent items within the carrier structure [tɒbeʃe] in non-presupposed and presupposed conditions of factive verb construction, pooled over 8 speakers. The values are given in semitones (reference: 100 Hz). Error bars represent standard error of the mean.

and information structure (non-presupposed, presupposed) as factors for factive verb construction. Huynh-Feldt corrected p-values are reported where appropriate.

For contrastive focus, the analysis revealed highly significant effects of minimal pair [$F(1,20) = 23.29$, $p < .001$, $\eta_p^2 = .538$], and information structure [$F(2,40) = 46.08$, $p < .001$, $\eta_p^2 = .697$], as well as a significant interaction effect between these factors [$F(2,40) = 13.11$, $p < .001$, $\eta_p^2 = .396$]. The Sidak post-hoc test showed that post-focal was significantly different from neutral ($p < .001$) and focal ($p < .001$). Taken as a whole, participants performed substantially worse in the post-focal condition than in the other two conditions, and the overall performance with the pair [tɒ́beʃ/tɒbéʃ] was significantly better than that with the pair [dídan/didán]. As indicated by the interaction between the two factors, the effect of focus was stronger on the identifications involving [tɒ́beʃ/tɒbéʃ].

As for factive verb construction, we again found significant effects of minimal pair [$F(1,20) = 30.95$, $p < .001$, $\eta_p^2 = .607$], and information structure [$F(1,20) = 89.40$, $p < .001$, $\eta_p^2 = .817$], as well as a significant interaction effect [$F(1,20) = 38.38$, $p < .001$, $\eta_p^2 = .657$]. The effects related to factive verb construction on the recognition are thus hardly distinguishable from the effects related to contrastive focus.
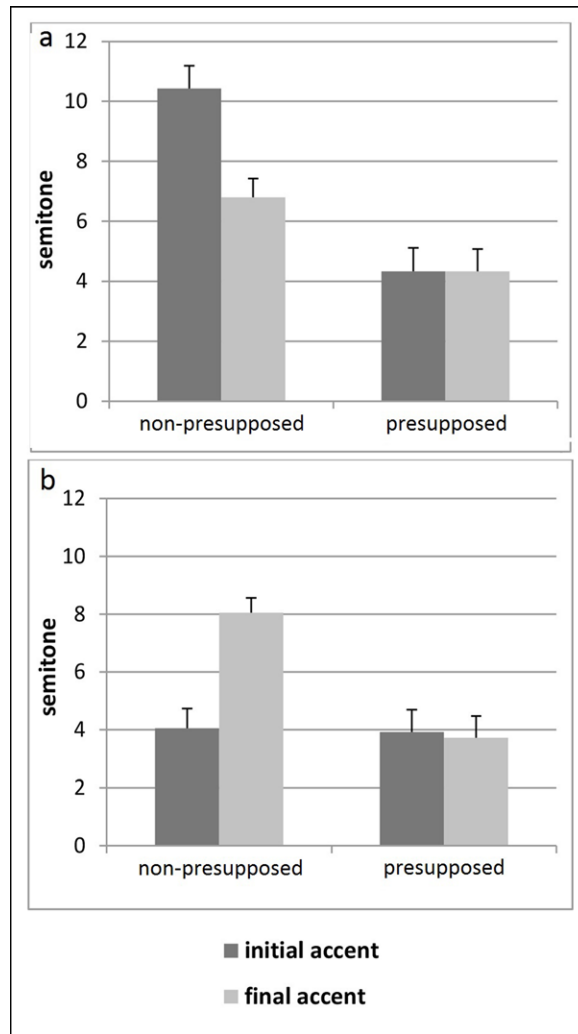
Figure 9: Mean F0 values of the 2nd half of the syllable [di] (panel (a)) and 2nd half of the syllable [da] (panel (b)) for initial accent and final accent items within the carrier structure [didaneʃ] in non-presupposed and presupposed conditions of factive verb construction, pooled over 8 speakers. The values are given in semitones (reference: 100 Hz). Error bars represent standard error of the mean.

Table 8: The experimental structure of the perception test.

Contrastive focus	8 (speakers) × 2 (minimal pairs) × 2 (accent positions) × 3 (information structure conditions) = 96 stimuli
Factive verb	8 (speakers) × 2 (minimal pairs) × 2 (accent positions) × 2 (information structure conditions) = 64 stimuli

Finally, we used one-sample t-tests to compare different performance conditions against baseline (i.e. $d' = 1.35$). According to the results, the identification of the pair [tʊ́beʃ/tʊbéʃ] was significantly better than the baseline in the neutral ($t_{20} = 5.41, p < .000$) and the focal conditions ($t_{20} = 4.63, p < .000$) related to contrastive focus, as well as in the non-presupposed condition related to factive verb construction ($t_{20} = 8.39, p < .000$). None of the other conditions were significantly above the baseline (see Table 9).

Our finding that the identification scores in the post-focal condition are significantly lower than those in the neutral and the focal conditions is consistent with the perception data presented in Abolhasanizadeh et al. (2012). The post-focal result in that study was however above some chance level in a way that our result is not. They obtained a score of 73% for the post-focal items, which was interpreted as being considerably
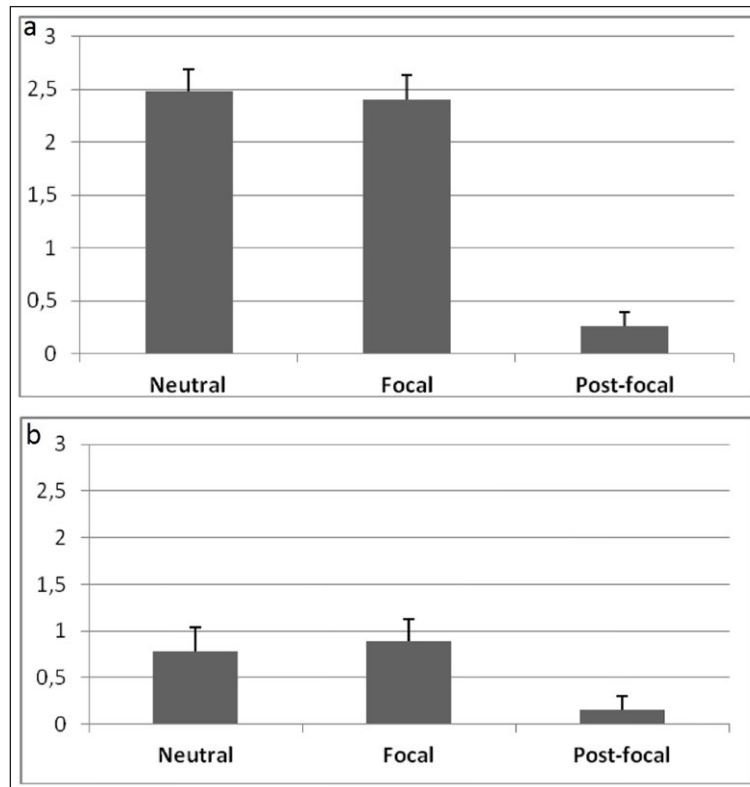
Figure 10: Mean *d'* scores in an identification task for the minimal pair [tʊ́beʃ/tʊbéʃ] (panel (a)), and the minimal pair [dídan/didán] (panel (b)), as obtained in neutral, focal and post-focal conditions related to contrastive focus, pooled over 21 listeners.
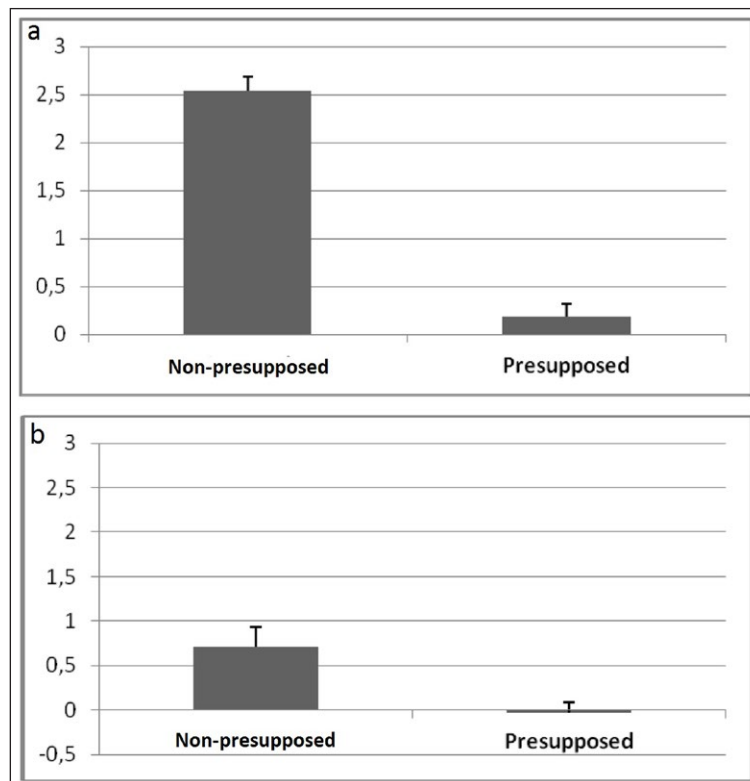

Figure 11: Mean *d'* scores in an identification task for the minimal pair [tʊ́beʃ/tʊbéʃ] (panel (a)), and the minimal pair [dídan/didán] (panel (b)), as obtained in non-presupposed and presupposed conditions related to factive verb construction, pooled over 21 listeners.

Table 9: Summary of the identification task.

Minimal pair	Information structure configuration	Experimental conditions	Successful Identification by the listeners?
[tóbeʃ/tɒbéʃ]	Contrastive focus	Neutral	Yes
		Focal	Yes
		Post-focal	No
	Factive verb construction	Non-presupposed	Yes
		Presupposed	No
[dídan/didán]	Contrastive focus	Neutral	No
		Focal	No
		Post-focal	No
	Factive verb construction	Non-presupposed	No
		Presupposed	No

above the chance level of 50%. In our data the mean scores (percentages correct) for the post-focal and the presupposed items are 54% for post-focal [tóbeʃ/tɒbéʃ], 53% for post-focal [dídan/didán], 53% for presupposed [tóbeʃ/tɒbéʃ] and 50% for presupposed [dídan/didán].

The results of Experiment I and Experiment II thus converge to suggest that accentual contrasts are neutralized in post-focal regions and presupposed dependent clauses. Given that there were no filler items in any of the two tasks, it is very likely that participants were aware of the minimal pair contrasts. It is remarkable that, despite the obviousness of the contrasts, speakers and listeners failed to differentiate between initial and final accent conditions in the post-focal and presupposed conditions (the unsuccessful identification of [dídan/didán] in all experimental conditions will be discussed in Subsection 4.1).

4 Discussion

4.1 The main result

The current study found that Persian word accent is deletable. Members of accentual minimal pairs become homophonous within post-focal regions and presupposed dependent clauses. This result is in accord with claims in the traditional Persian grammars that accent contrasts are neutralized in such contexts.

Our finding disconfirms the results of Abolhasanizadeh et al. (2012). Because their participants recognized the members of similar minimal pairs above chance, they tentatively concluded that constituents retain the accent after the focus. The explanation of the discrepancy most probably lies in the method of data collection. Instead of a contextualized elicitation task, they adopted a reading task in which the focused constituents were presented in bold print. The lack of an appropriate context may have led some speakers to pronounce the target items merely with more or less emphasis rather than with focus prosody. Note that focus, a grammatical concept, need not co-occur with (paralinguistic) emphasis. We return to this point below.

There is one finding in our data that was unexpected. While both minimal pairs ([tóbeʃ/tɒbéʃ] and [dídan/didán]) were found to be deaccented in the post-focal condition (related to contrastive focus) and in the presupposed condition (in dependent clauses of factive verbs), the overall recognition score was significantly lower for the pair [dídan/didán]. In fact, it did not even reach the baseline level. The question is why listeners performed poorly with the accented forms of this pair in both the neutral/focal

conditions and the non-presupposed condition. Participants cannot have had any difficulty identifying accent locations, given that Persian speakers are widely exposed to the contrastive function of accent and will immediately spot incorrect accent placements. This position is supported by high recognition scores for the members of the minimal pair [tɒ́beʃ/tɒbéʃ]. We suggest that participants must have had some difficulties with the association of target items with the response buttons, despite the fact that they all indicated that they had mastered the association after the trial stimuli. Their confusion may have arisen because the colloquial minimal pair [dídan] 'they saw' vs. [didán] 'they have seen' involves a grammatical contrast between past and perfect forms of the same verbal stem, for which speakers may have a less clearly defined intuition than for lexical contrasts. Recall that these items appeared on the screen in their literary forms ([didand] vs. [dideand]). The members of the other minimal pair, by contrast, represented different lexemes with clearly different meanings. Moreover, there is no
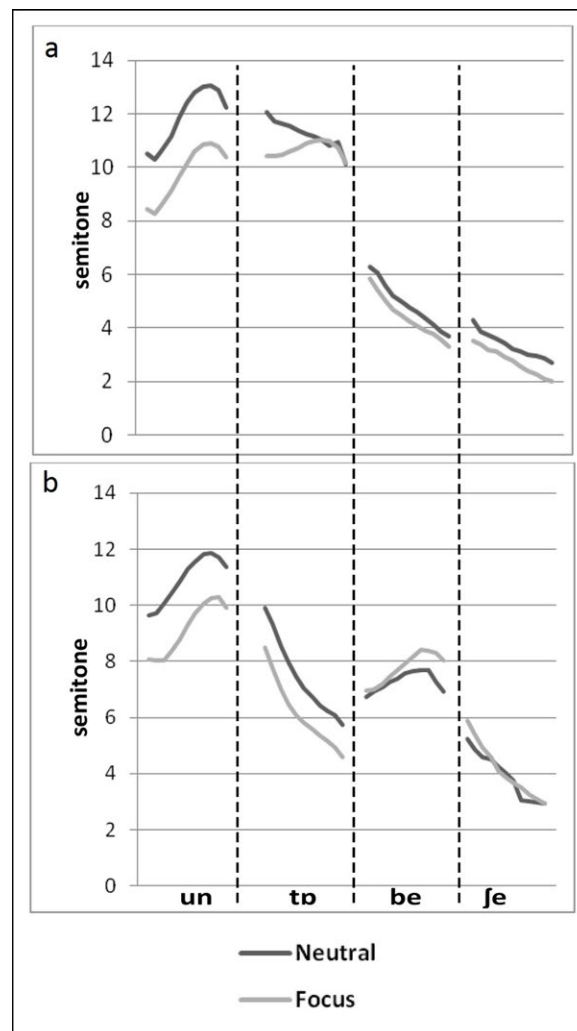
Figure 12: Mean F0 contours for [ún tɒ́beʃe] (panel (a)) and [ún tɒbéʃe] (panel (b)) on a normalized time scale for neutral and focal pronunciations separately, pooled over 8 speakers. The values are given in semitones (reference: 100 Hz).
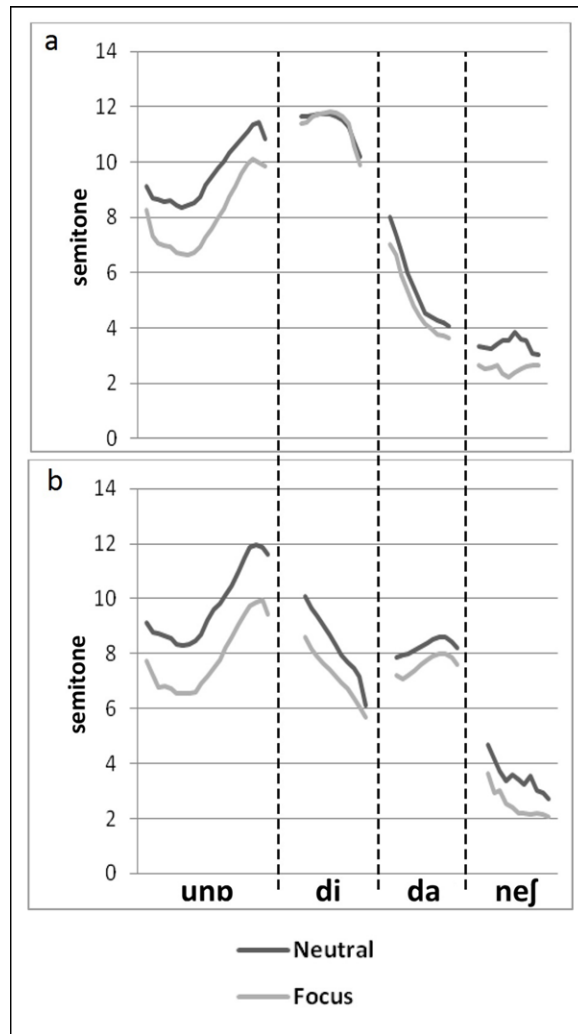
Figure 13: Mean F0 contours for [unɒ́ dídaneʃ] (panel (a)) and [unɒ́ didáneʃ] (panel (b)) on a normalized time scale for neutral and focal pronunciations separately, pooled over 8 speakers. The values are given in semitones (reference: 100 Hz).

difference between the colloquial and literary realizations of these items, unlike the situation for [dídan/didán].[26]

4.2 Focus accent vs. neutral accent

One observation in our production data (Experiment I) concerns the question of whether accent for focus is phonologically different from neutral accent in Persian. As noted in Subsection 1.3, previous experimental studies have shown that focused constituents are realized with a higher F0 (Sadat-Tehrani 2007; Taheri & Xu 2012; Hosseini 2013). It has been suggested that the extra high on (contrastively) focused elements might be phonological rather than phonetic (Scarborough 2007). To verify this idea, we show contrastive focus contours against neutral contours for each member of the two minimal pairs separately in Figure 12, which gives time-normalized mean curves for [tɒ́beʃ] (panel

---

[26] The difficulty with the identification of the verbal pair [dídan/didán] 'they saw/they have seen' might be generalized to other past vs. perfect accentual contrasts, as suggested by informal observations. In a small-scale experiment, listeners showed difficulty in distinguishing between members of the two minimal pairs [ráftim/raftím] 'we went/we have gone' and [umádin/umadín] 'you came/you have come'.

(a)) and [tɒbéʃ] (panel (b)), while Figure 13 does the same for [dídan] (panel (a)) and [didán] (panel (b)).

Our data do not confirm the claim that the accent for focus is consistently higher than the neutral accent, suggesting that F0-raising is not a consistent mechanism to mark focus in Persian. Perceptual data reported by Taheri et al. (2014) support this view of F0-raising for focus as a paralinguistic and optional speaker strategy, though these authors did not provide any discussion. Taheri et al. (2014) conducted a perception task in which Persian listeners were asked to identify the position of the focused item in a five-word sentence. While the mean and maximum F0 of all the focused constituents in their corpus were substantially higher than that of their neutral counterparts, the identification scores were lower for the sentence-final items (59%) compared to the non-final items (79%). This suggests that absence of PFC, which is not available in sentence-final position, inhibited the recognition of the focus for sentence-final target items. Since identification errors in final position were biased towards neutral focus (73%), this result supports the conclusion that there is no phonological difference between the accent for focus and neutral accent in Persian and that the two sentences have the same phonological structure.

To sum up, focus is marked by post-focal compression, such that elements with neutral focus and narrow focus are homophonous (see also the discussion in Subsection 4.3).

A final observation in Figures 12 and 13 is that the pitch register of utterances with narrow focus was unexpectedly lower than in neutral utterances, as is visible in Figure 13 (panel (b)), where the focus contour appears to be shifted downward with respect to the neutral contour. This use of a lower register in focus utterances may be interpreted in terms of Ohala's (1983) Frequency Code (FC), which is based on the fact that smaller larynxes produce higher fundamental frequency than larger ones, causing high pitch to sound uncertain or submissive and low pitch to sound authoritative and assertive. Conceivably, speakers used a lower register to sound more authoritative in contrastively focused utterances compared to neutral utterances. This observation is consistent with the findings of Kügler & Genzel (2012), who show that Akan's H and L tones have lower pitch under corrective focus.[27]

4.3 Focus vs. emphasis

The reason why the earlier literature has frequently reported extra high F0 for Persian focus may be due to experimental procedures that confound the effect of focus with that of (paralinguistic) emphasis. In such procedures, participants are instructed in a way that may give them the impression that the task is about emphasizing words, i.e. pronouncing words with a great energy, resulting in higher F0, higher intensity and greater duration. In many cases, experimenters present participants with bold or underlined letters in target items. Moreover, it is frequently the case that during the training sessions, the experimenters use the Arabic loanword [taʔkid] 'emphasis' to refer to focus, as for example in the case of Abolhasanizadeh et al. (2012) (Mahmood Bijankhan, p.c.). As we noted in Subsection 2.3, we did not highlight target items in our experiment, instead relying on context sentences to elicit the intended focus structure. In a series of pilot experiments, in which we did use bold print for target items, we found that some participants pronounced them at a higher-than-average pitch. To illustrate the distinction between the two procedures, we present four realizations of the target sentence in (12) in Figure 14.

(12)   nili = o   landan   did = im.

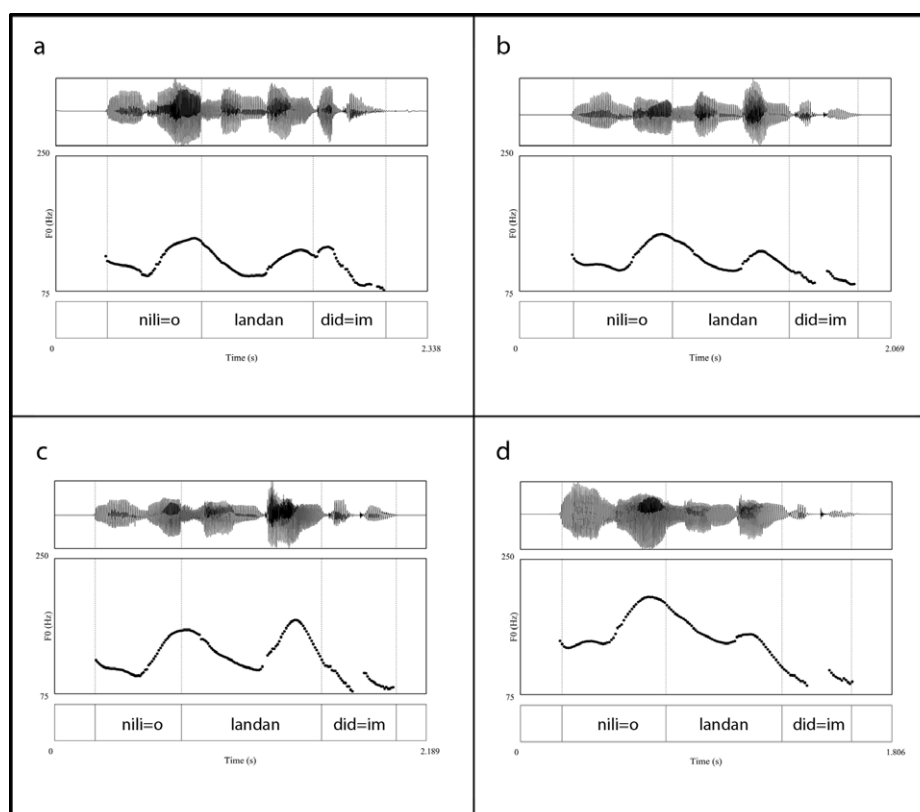[27] We thank an anonymous reviewer for this reference.

Figure 14: F0 contours of four realizations of the sentence [nilio landan didim] 'We saw Nili in London' as variations of focus and/or emphasis, spoken by a male speaker. Panel (a) has focus on the whole sentence and no particular emphasis on any word. Panel (b) has focus on [landan] and no particular emphasis on any word. Panel (c) has focus and emphasis on [landan]. Panel (d) has focus on [landan] and emphasis on [nili].

Nili = OM   London   saw = 1PL
'We saw Nili in London.'

The realization of (12) shown in panel (a) of Figure 14 was produced in response to *What happened?*, while that in panel (b) was spoken in response to *Where did you see Nili?* and has narrow focus on [landan] 'London'. Neither was spoken with particular emphasis. To illustrate the prosodic effect of emphasis, panels (c) and (d) give realizations of (12) with focus for [landan], obtained using our contextual elicitation procedure. In panel (c), we observe F0-raising on the focused item compared with the pronunciation in panel (b). The contour in panel (d) shows F0-raising is not necessarily applied to the focused item, since here non-focused [nili] has a significantly raised F0 peak as compared to the other contours in Figure 14 and to the focused item in the same utterance. These data suggest that the realization of focus and the application of emphasis are independent. Focus is indicated by PFC, regardless of the presence of emphatic pronunciation before the focus.

4.4 *Prosodic vs. morphosyntactic expression of focus*

The prompts and stimuli in our experiments illustrate *prosodic focus,* meaning that there is no marking of focus other than through prosody. Karimi (2005) in particular has argued that there is a morphosyntactic strategy to express (contrastive) focus through the use of exclusive focus adverbs like [faɢat] 'only', which in the usual cases appears to the left of
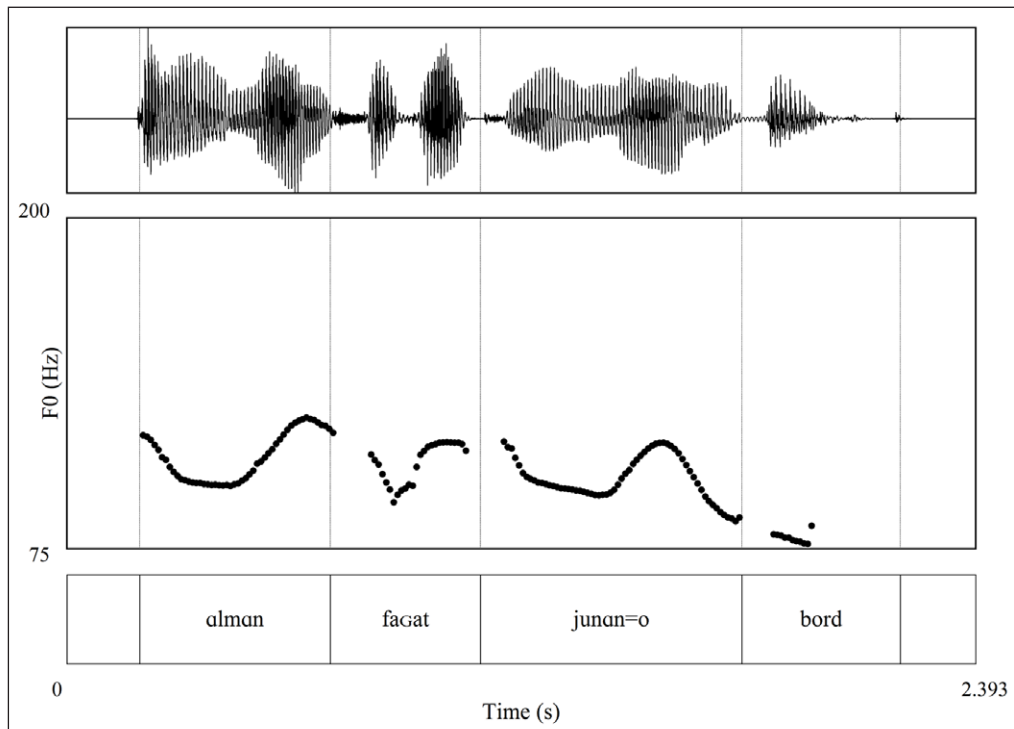
Figure 15: F0 contour of [ɒlmɔ́n faɢát junɔ́n=o bord] 'It was only Greece that Germany beat', spoken by a male speaker.

the focus constituent, as in (13), where the object [junɒn] 'Greece' represents the focus constituent.

(13)   ɒlmɒn   faɢat   [junɒn]$_{FOC}$ = o   bord.
        Germany   only   Greece = OM   beat.3SG
        'It was only Greece that Germany beat.'

Karimi claims that contrastive focus is realized either morphosyntactically or prosodically, which suggests that the focus for [junɒn] in (14) can either be expressed through deaccentuation of the verb [bord], as in (14a), or by attaching the focus adverb to the object, as in (14b).

(14)   ɒlmɒn   junɒn = o   bord.
            Germany   Greece = OM   beat.3SG
            'Germany beat Greece.'
        a.   ɒlmɔ́n [junɔ́n]$_{FOC}$ = o bord.
        b.   ɒlmɒn faɢat [junɒn]$_{FOC}$ = o bord.

However, the use of a focus adverb like [faɢat] 'only' in no way absolves the speaker from employing prosodic focus marking in the form of post-focus deaccentuation, as shown in (15) with its F0 contour depicted in Figure 15 (cf. panel (c) in Figure 1 which gives the F0 contour of (14a)). Prosodic focus marking is thus obligatory and independent of the morphosyntactic focus marker.

(15)   ɒlmɔ́n faɢát [junɔ́n]$_{FOC}$ = o bord.

Moreover, prosodic focus may be the only way in which the scope of the adverb is signalled, as shown in (16). The focus adverb here takes scope over the preceding subject, since the elements after the focus adverb are deaccented.[28],[29]

(16)   [ɒlmɑ́n]_FOC faɢat junɒn = o bord.
        'It was only Germany that beat Greece.'

These observations strongly suggest that adverbs like [faɢat] cannot express focus independently of the prosodic focus marking.[30] This finding, along with our observations in Subsection 1.3 regarding strict morphosyntactic constraints on the realization of prosodic focus, highlight the close link between prosodic focus and morphosyntactic structure in Persian.

5 Conclusion

This study aimed to establish whether Persian word accents are deleted in post-focal regions and presupposed embedded clauses, such that the members of accentual minimal pairs become homophonous. The results of two experiments confirmed that accents are in fact deleted. A production experiment showed low F0 plateaus on the post-focal and presupposed constituents, while perception experiment showed that deaccented members of minimal pairs are not recognized above the just-noticeable-difference (JND) baseline.

The finding that Persian deaccents after the focus concurs with Xu et al.'s (2012) suggestion that post-focal compression of the pitch range (PFC) may be an areal characteristic extending from Western Europe to central Asia. It is to be noted, however, that languages have different ways of creating the effect of PFC. Persian, having no word stress (Abolhasanizadeh et al. 2012), is a clear case of a language that neutralizes the accent contrast. English and Dutch equally deaccent words in the post-focal region (though see Prom-On et al. 2009), but because distinctions between primary stress locations, probably including those between primary stress and secondary stress, may be preserved, they are not neutralizing in the way Persian is. Similarly, lexical tones are preserved under PFC in Mandarin (Prom-On et al. 2009), while Japanese leaves word accentual distinction intact, because PFC is again restricted to pitch range compression, possibly together with a boosting of the focus without deletion of tones (Pierrehumbert & Beckman 1988; Ito 2002). Inasmuch as there would appear to be a tendency for lexical prosodic structure to be preserved and post-lexical structure to be deleted, the finding by Rahmani et al. (2015) that the Persian tone is assigned post-lexically is consistent with our finding that it can be deleted so as to neutralize contrasts.

A final remark here is that the morphosyntactic nature of accent distribution in sentences does not bear on the generalization that post-lexical phonology is governed by the prosodic hierarchy (Nespor & Vogel 1986; Selkirk & Lee 2015). The Persian accent is a morpheme that, like other morphemes, is incorporated in the morphosyntactic representation, to which phonological rules may apply once it is included in a prosodic

---

[28] Note that the focus adverb is itself deaccented in (16). Otherwise, (16) would be interpreted as having narrow focus on the adverb, a situation that might arise in cases of metalinguistic correction.

[29] Under specific discourse conditions, the adverb in (16) may be interpreted as taking scope over the object (or the object + verb combination), in particular in cases in which the scope is repeated information. Thus, in reply to *Who beat only Greece?*, [ɒlmɑ́n faɢat junɒn = o bord] is interpreted to mean that *Germany beat only Greece*. Informal observations suggest that prosodic phrasing, i.e. durational cues, may be a factor in indicating the scope of the adverb in such contexts (cf. Beaver et al. 2007).

[30] Our observations on focus adverbs can be extended to some other morphosyntactic focus markers discussed in Karimi (2005), which we have ignored in the main text for the sake of brevity. In particular, as Karimi (2005) claims, a contrastively focused element may be marked by means of movement to sentence-initial position. Here, it should be noted that placing a non-sentence-initial constituent at the front of the sentence does not mark that constituent as focus unless all the following elements are deaccented.

constituent structure. The same is true for English accents, as exemplified by predicate deaccentuation as in *The SUN is shining* or deaccentuation of second constituents in noun compounds, as in *SUNshine*. The difference with Persian is that English happens to also have a phonological accent deletion rule, the Rhythm Rule as applicable inside the phonological phrase (cf. *DUNdee MARmelade*, as contrasting with *MARmelade from DunDEE*; Gussenhoven 2011).

Additional file

The additional file for this article can be found as follows:

- **Appendix 1:** Corpus belonging to Experiment I. DOI: https://doi.org/10.5334/gjgl.328.s1

Abbreviations

1/2/3 = person, ART = article, COP = copula, DER = derivational, DUR = durative, FOC = focus, OM = object marker, PP = past participle, POSS = possessor, SG/PL = singular/plural, F0 = fundamental frequency, JND = just-noticeable-difference, PFC = post focus compression, SD = standard deviation, ST = semitone

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Abolhasanizadeh, Vahideh, Mahmood Bijankhan & Carlos Gussenhoven. 2012. The Persian pitch accent and its retention after the focus. *Lingua* 122(13). 1380–1394. DOI: https://doi.org/10.1016/j.lingua.2012.06.002

Beaver, David, Brady Clark, Edward Flemming, T. Florian Jaeger & Maria Wolters. 2007. When semantics meets phonetics: Acoustical studies of second-occurrence focus. *Language* 83(2). 245–276. DOI: https://doi.org/10.1353/lan.2007.0053

Beckman, Mary E. 1986. *Stress and non-stress accent*. Dordrecht: Foris. DOI: https://doi.org/10.1515/9783110874020

Bijankhan, Mahmood. 2005. *Vajšenasi: Næzæriye-ye behinegi* [Phonology: Optimality Theory]. Tehran: Samt.

Boersma, Paul & David Weenink. 2014. Praat: Doing phonetics by computer [computer program]. http://www.praat.org.

Crystal, David. 2008. A *dictionary of linguistics and phonetics*. Malden, MA: Blackwell. DOI: https://doi.org/10.1002/9781444302776

Dupoux, Emmanuel, Sharon Peperkamp & Núria Sebastián-Gallés. 2001. A robust method to study stress "deafness". *The Journal of the Acoustical Society of America* 110(3). 1606–1618. DOI: https://doi.org/10.1121/1.1380437

Durrant, John D., Jean H. Lovrinic. 1977. *Bases of hearing science*. Baltimore, MD: Williams & Wilkins. DOI: https://doi.org/10.1121/1.391810

Elordieta, Gorka & Aritz Irurtzun. 2010. The relationship between meaning and intonation in non-exhaustive answers: Evidence from Basque. *The Linguistic Review* 27(3). 261–291. DOI: https://doi.org/10.1515/tlir.2010.010

Eslami, Moharram. 2000. *Šenaxt-e næva-ye goftar-e zæban-e Farsi væ karbord-e an dær bazsazi væ bazšenasi-ye rayane'i-ye goftar* [The prosody of the Persian language and its application in computer-aided speech recognition]. Tehran: University of Tehran dissertation.

Ferguson, Charles A. 1957. Word stress in Persian. *Language* 33(2). 123–135. DOI: https://doi.org/10.2307/410724

Ghaniabadi, Saeed, Jila Ghomeshi & Nima Sadat-Tehrani. 2006. Reduplication in Persian: A morphological doubling approach. In Claire Gurski & Milica Radisic (eds.), *Proceedings of the 2006 Annual Conference of Canadian Linguistic Association*, 1–15. Toronto: University of Toronto.

Ghomeshi, Jila. 1996. *Projection and inflection: A study of Persian phrase structure*. Toronto: University of Toronto dissertation.

Gussenhoven, Carlos. 2008. Notions and subnotions in information structure. *Acta Linguistica Hungarica* 55(3–4). 381–395. DOI: https://doi.org/10.1556/ALing.55.2008.3-4.11

Gussenhoven, Carlos. 2011. Sentential prominence in English. In Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 5. 2780–2806. Malden, MA & Oxford: Wiley Blackwell. DOI: https://doi.org/10.1002/9781444335262.wbctp0116

Hamlaoui, Fatima, Sascha Coridun & Caroline Féry. 2012. Expression prosodique du focus et du donné au sein des groupes nominaux [NA] du français. In Franck Neveu, Valelia Muni Toke, Peter Blumenthal, Thomas Klinger, Pierluigi Ligas, Sophie Prévost & Sandra TesonBonnard (eds.), *Actes du 3e Congrès Mondial de Linguistique Française*, 1505–1518. Paris: EDP Sciences.

Hooper, Joan B. & Sandra A. Thompson. 1973. On the applicability of root transformations. *Linguistic inquiry* 4(4). 465–497.

Hosseini, Ayat. 2014. *The phonology and phonetics of prosodic prominence in Persian*. Tokyo: Tokyo University of Foreign Studies dissertation.

Ito, Kiwako. 2002. *The interaction of focus and lexical pitch accent in speech production and dialogue comprehension: Evidence from Japanese and Basque*. Champaign, IL: University of Illinois at Urbana-Champaign dissertation.

Kahn, Daniel. 1976. *Syllable-based generalizations in English phonology*. Cambridge, MA: MIT unpublished dissertation. [distributed by Indiana University Linguistics Club, Bloomington, IN].

Kahnemuyipour, Arsalan. 2003. Syntactic categories and Persian stress. *Natural Language & Linguistic Theory* 21(2). 333–379. DOI: https://doi.org/10.1023/A:1023330609827

Karimi, Simin. 2005. *A Minimalist approach to scrambling: Evidence from Persian*. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110199796

Kiparsky, Paul & Carol Kiparsky. 1970. Fact. In Manfred Bierwisch & Karl Erich Heidolph (eds.), *Progress in Linguistics,* 143–173. The Hague: Mouton.

Kügler, Frank & Susanne Genzel. 2012. On the prosodic expression of pragmatic prominence: The case of pitch register lowering in Akan. *Language and speech* 55(3). 331–359. DOI: https://doi.org/10.1177/0023830911422182

Lazard, Gilbert. 1992. *A grammar of contemporary Persian*. Costa Mesa, CA: Mazda Publishers.

Macmillan, Neil A. & C. Douglas Creelman. 1991. *Detection theory: A user's guide.* Cambridge: Cambridge University Press.

Macmillan, Neil A. & Howard L. Kaplan. 1985. Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological bulletin* 98(1). 185–199. DOI: https://doi.org/10.1037/0033-2909.98.1.185

Mahjani, Behzad. 2003. *An instrumental study of prosodic features and intonation in modern Farsi (Persian).* Edinburgh: University of Edinburgh MA thesis.

Mahootian, Shahrzad. 1997. *Persian descriptive grammar.* London: Routledge.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology.* Dordecht: Foris.

Ohala, John J. 1983. Cross-language use of pitch: An ethological view. *Phonetica* 40(1). 1–18. DOI: https://doi.org/10.1159/000261678

Pierrehumbert, Janet & Mary Beckman. 1988. *Japanese tone structure.* Cambridge, MA: MIT Press. DOI: https://doi.org/10.1017/S0022226700014286

Prom-On, Santitham, Yi Xu & Bundit Thipakorn. 2009. Modeling tone and intonation in Mandarin and English as a process of target approximation. *The Journal of the Acoustical Society of America* 125(1). 405–424. DOI: https://doi.org/10.1121/1.3037222

Rahmani, Hamed, Toni Rietveld & Carlos Gussenhoven. 2015. Stress "deafness" reveals absence of lexical marking of stress or tone in the adult grammar. *PloS One* 10(12). DOI: https://doi.org/10.1371/journal.pone.0143968

Ross, John Robert. 1973. Slifting. In Maurice Gross, Morris Halle & Marcel-Paul Schützenberger (eds.), *The formal analysis of natural languages. Proceedings of the First International Conference,* 133–169. The Hague: Mouton.

Sadat-Tehrani, Nima. 2007. *The intonational grammar of Persian.* Winnipeg: University of Manitoba dissertation.

Scarborough, Rebecca. 2007. The intonation of focus in Farsi. *UCLA Working Papers in Phonetics* 105. 19–34.

Selkirk, Elisabeth & Seunghun J. Lee. 2015. Constituency in sentence phonology: An introduction. *Phonology* 32(1). 1–18. DOI: https://doi.org/10.1017/S0952675715000020

Simons, Mandy. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117(6). 1034–1056. DOI: https://doi.org/10.1016/j.lingua.2006.05.006

Taheri-Ardali, Mortaza, Hamed Rahmani & Yi Xu. 2014. The Perception of prosodic focus in Persian. In Nick Campbell, Dafydd Gibbon & Daniel Hirst (eds.), *Proceedings of the 7th International Conference on Speech Prosody,* 515–519. Dublin: Trinity College.

Taheri-Ardali, Mortaza & Yi Xu. 2012. Phonetic realization of prosodic focus in Persian. In Qiuwu Ma, Hongwei Ding & Daniel Hirst (eds.), *Proceedings of the 6th International Conference on Speech Prosody,* 326–329. Shanghai: Tongji University Press.

Vahidian-Kamyar, Taghi. 2001. *Næva-ye goftar dær Farsi* [Melody of speech in Persian]. Mashhad: Ferdowsi University Press.

Windfuhr, Gernot. 1979. *Persian grammar: History and state of its study.* The Hague/Paris/ New York: Mouton. DOI: https://doi.org/10.1515/9783110800425

Xu, Yi, Szu-Wei Chen & Bei Wang. 2012. Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family? *The Linguistic Review* 29(1). 131–147. DOI: https://doi.org/10.1515/tlr-2012-0006

Zwicky, Arnold M. & Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English n't. *Language* 59. 502–513. DOI: https://doi.org/10.2307/413900