Choosing the optimal pitch accent location in Dutch by Chinese learners and native listeners

He, X.; Heuven, V.J. van; Gussenhoven, C.; Dziubalska-Kolaczyk, K; Wrembel, M.; Kul, M.

Choosing the optimal pitch accent location in Dutch by Chinese learners and native listeners

Xuliang He[1,2]*, Vincent J. van Heuven*[3] *and Carlos Gussenhoven*[2]

[1]Nantong University, [2]Centre for Language and Communication, Radboud University Nijmegen
[3]Leiden University Centre for Linguistics
{x.he, c.gussenhoven}@let.ru.nl, v.j.j.p.van.heuven@hum.leidenuniv.nl

ABSTRACT

A perception experiment was conducted to study how well Chinese learners of Dutch identify the correct accentuation patterns in six categories of Dutch sentences. Thirty-six stimuli (6 sentences x 6 categories) were presented to 20 Dutch native listeners (NLD) and 20 Chinese learners of Dutch as a second language (CLD). In a forced-choice task, listeners had to decide which of two versions of each sentence was pronounced with optimal prosody, and to indicate how confident they were about their choice on a five-point scale. Per test item only one accent placement was prosodically optimal ('correct') as determined prior to the test by a panel of Dutch intonologists.

NLD correctness scores were significantly higher than those by the Chinese learners with high proficiency in Dutch (CLD-H), and the correctness scores of the latter were also significantly higher than those obtained for the low-proficiency learners (CLD-L). Along with the correctness scores, confidence ratings decreased from NLD to CLD-H to CLD-L. The results show that the different categories of accent placement do not present the same degree of difficulty to the Chinese learners.

Keywords: perception, accent, Dutch, Chinese learners of Dutch

1. INTRODUCTION

Speakers of non-native languages (L2 speakers) often have difficulties in producing acceptable stress and accentuation patterns, in part depending on the difference between L1 and L2. Incorrect prominence patterns often persist despite long exposure to the L2. It would appear that such 'stress deafness' is not merely the result of major stress typology differences (Gut, Trouvain and Barry 2007). One situation in which it has been observed is that of Chinese learners of English (Trouvain and Gut 2007). Germanic languages generally have sentence prosodic patterns that are the result of the inherent stress patterns of words and the rather complex rules for the placement of pitch accents on a subset of the stressed syllables. Variation in the distribution of pitch accents is strongly context-dependent, but even in reading tasks with isolated sentences, differences between native speakers and Chinese L2 speakers are striking (Chen Hua 2008). The question we attempt to address in this investigation is first, whether the apparent difficulties Chinese learners of Dutch (CLD) have with the correct placement of prominences in Dutch sentences is due to their inability to produce those patterns or rather in their ignorance of what an acceptable pronunciation of an isolated sentence is. The question, therefore, is whether they can recognize the correct location of sentence accents in a listening task. Additionally, given that the explanation for the presence or absence of a pitch accent on a given syllable may vary from morphology to information structure, our interest is in whether the acceptability of some accent placements are easier to establish by Chinese learners of Dutch than other accent placements. To this end, we have classified the accent placements according to the linguistic generalization that lies at their basis.

Accent is a place marker in the phonological structure where tones, known as pitch accents, are to be inserted (Goldsmith 1976, Hyman 1978). Pitch accents in a Dutch sentence are determined by lexical, phonological and morphological information, as well as by semantic and pragmatic factors (Gussenhoven ms). We chose six categories of accentuation problems in sentences for our participants to judge in the experiment.

The first category concerns primary word stress, which in Dutch falls on the antepenult, the penult, or the final syllable of the word if the penult is open, and on the penult or the final syllable if the penult is closed

(Gussenhoven 1999, Booij 1999). In the word *ooievaar* 'stork', for instance, in which the first and the last syllables are stressed, the primary stress falls on the antepenult *oo* and not on the last, *vaar*. Because final VVC syllables regularly take the primary stress, exceptional words like *ooievaar*, *Spanjaard* 'Spaniard', *olifant* 'elephant', may present difficulties to foreign learners, even though such words are both common and frequent. In the case of *ooievaar* (1a) is the correct reading.

(1) a. Er staat een OOievaar in de wei.
b. Er staat een ooieVAAR in de wei.
'There's a stork in the meadow.'

(2) a. Ze hebben een grote GROENTEtuin.
b. Ze hebben een grote groenteTUIN.
'They have a large kitchen garden.'

The second problem category comprises compound nouns. Of the two constituents in the compounds, the second loses its pitch accent. For example, the primary stress of *groentetuin* 'kitchen garden' is on the first component *groente* rather on the second *tuin*, which means that reading (2a) is correct.

The third problem is that of phrasal proper names. People's names generally consist of the given name, the surname and an optional surname prefix. The prefix is a function word (or multiple function words), which remains unstressed, and is considered a part of the surname. Some surnames, particularly those that were adopted as a result of the compulsory registration of the Dutch population during the French occupation of 1806-1813, are etymologically phrases, like *Vroegindewei*, which goes back to *Vroeg in de wei* 'early in the meadow'. The primary stress of a phrasal surname generally falls on the first syllable, but on the last if the same words are used as a common phrase. That is, (3a) is the appropriate reading.

(3) a. Dit is mevrouw VROEGindewei.
b. Dit is mevrouw vroegindeWEI.
'This is Mrs Vroegindewei.'

(4) a. Ze hebben een aantal nieuwe WERken aangekocht.
b. Ze hebben een aantal nieuwe WERken AANgekocht.
'They've acquired a number of new works of art.'

Gussenhoven (1983) made a distinction between eventive and non-eventive sentences, where eventive sentences involve the reporting of a change in the world. Following Schmerling (1976), the Sentence Accent Assignment Rule (SAAR) says that in eventive sentences, predicates lose their pitch accent if they are adjacent to one of their arguments. For instance, if a single accent falls upon *dogs* in the sentence *[DOGS must be carried]*$_{\text{EVENTIVE}}$, all relevant people will have to carry a dog. A non-eventive sentence would retain accents on both the argument *dogs* and the predicate *must be carried*, in which 'contingency' reading only those people who happen to have a dog are obligated to carry it (Ladd 2008). The fourth and the fifth categories are eventive and non-eventive sentences, respectively. In (4a), the argument of the verb *WERken* 'works' is accented and the verb *aangekocht* 'acquired' is deaccented. That is the reason why (4a), an eventive sentence, is correct. However, the (a) version is the correct reading in the case of (5), because *schade* 'damage' only exists potentially, so that the verb *MELden* 'report' is accented.

(5) a. U wordt verzocht eventuele SCHAde te MELden.
b. U wordt verzocht eventuele SCHAde te melden.
'You are requested to report any damage.'

Deaccenting for 'givenness' and accenting for 'newness' is illustrated in (6), where *jas* 'coat' is given information in the first sentence, and *zwarte* 'black' is the new information. That is, (6a) is correct, while (6b) is not.

(6) a. Ik heb wel een mooie bruine jas gezien. Maar ik zocht eigenlijk een ZWARTE jas.
b. Ik heb wel een mooie bruine jas gezien. Maar ik zocht eigenlijk een zwarte JAS.
'I did see a nice brown coat. But I was really looking for a black coat'

Mandarin Chinese has two types of accentuation: grammatical and logical (Gao 1984). Grammatical accents are determined by the structure of syntactic phrases and logical accents depend upon the meanings speakers intend to express. Chinese is a tone language, while Dutch is an intonation-only language. Inevitably, Chinese speakers of Dutch are influenced, to greater or lesser extent, by their L1 Chinese when they speak

Dutch. We know from the field of second language acquisition that the majority of second-language learners cannot acquire native-like oral ability, but less is known about the extent to which non-native speakers acquire the prosodic knowledge of the target language even in a situation in which they cannot produce it correctly. The answer to that question is important, because the cause of any mispronunciations will need to be attributed either to a lack of knowledge of the L1's prosodic structure or to an inability to pronounce such structures. The present perception experiment was conducted to study how well Chinese speakers of Dutch identify the correct accentuation pattern in the six categories of Dutch sentences. The first question we addressed is whether Chinese learners of Dutch are less often correct and less confident in their judgment of the appropriateness of accent patterns in Dutch than native Dutch listeners, and if their performance varies over the six accent placement categories. The second question is whether and to what extent Chinese speakers with higher overall proficiency in Dutch also do better on the accent judgment task than their counterparts with lower proficiency in Dutch.

2. METHOD

2.1. Materials

We obtained six categories of Dutch sentences with different accent patterns in each category (see section 1). Each category contains six sentences, so we get 36 sentences altogether in the corpus.

The recordings were made in a sound-treated booth at Radboud University Nijmegen. Each sentence was read with correct and incorrect accentuation by the third author or by a female Dutch phonologist. Two sets of stimuli included 72 [6 sentences x 2 (correct and incorrect) x 6 categories = 72] different stimuli with each sentence read with correct and incorrect accentuation by different speakers. The two sets had the same random order of sentences with the complementary stimuli, one mirrored the other. During the recordings, the third author checked to make sure the target words were read with proper accentuation. And we chose the best token of each pair of recordings as stimuli presented to the listeners.

2.2. Participants

A group of 20 native and 20 Chinese speakers of Dutch participated in the perception test. All Chinese participants (3 male, 17 female) were from the northern part of China, aged from 17 to 53. At the time of the experiment they had lived in the Netherlands for periods between three months and 22 years. The two subjects who had lived in the country for three months had studied Dutch in China for more than two years. This means that all Chinese participants had had sufficient exposure to the language to be able to do the experiment. The Dutch participants (4 male, 16 female) were self-declared native speakers of standard Dutch, aged between 18 to 54 years old. They were divided into two groups to do the forced-choice task. Ten Chinese subjects and ten Dutch subjects listened to stimuli in Set I and the other half in Set II.

The twenty Chinese subjects were asked to read a text of 42 Dutch sentences (our Production Experiment 3), and the Chinese group of speakers was divided into a higher ('Chinese Listeners of Dutch Higher': CLD-H) and a lower (CLD-L) subgroup on the basis of each subject's mean score over their segmental and prosodic proficiency scores judged by three experts. Pearson's correlations between the two scores between experts are $r_{pro1, pro2} = 0.97$, $r_{pro1, pro3} = 0.85$, $r_{pro2, pro3} = 0.88$; $r_{seg1, seg2} = 0.82$, $r_{seg1, seg3} = 0.76$, $r_{seg2, seg3} = 0.97$. This means three experts were highly constant to each other on the judgment of subjects' proficiency.

2.3. Procedure

Set I or Set II of stimuli was presented in individual sessions on a computer screen. Each participant was asked to listen to the stimuli and judge whether the reading was correct or not. The listener wore a GH632 headphone. Subjects first listened to 8 trial stimuli and then proceeded to the experiment proper. They could replay any stimulus before finalizing a judgment. The correct and incorrect tokens were randomized across categories. The listeners first clicked either of the two buttons marked 'correct' and 'incorrect', and then indicated his confidence in the judgment on a scale from 1 (poor confidence) to 5 (high confidence).

3. DATA ANALYSIS AND DISCUSSION

3.1. Analysis and discussion of correctness scores

We conducted a repeated measures Analysis of Variance (RM-ANOVA) using accent type (A: word stress, B: compound, C: proper names, D: eventive sentences, E: Non-eventive sentences, F: focus) as within-subjects variables and the different language groups as a between-subjects factor. The results show that correctness scores (Table 1) are significantly different ($F [5, 190] = 4.9$, $p < .05$) across accent types. Pairwise comparisons reveal that mean differences between types A and D, and between D and F are significant ($p = .014$, $p = .002$, respectively). There is no significant interaction between accent type and language group ($F [5, 190] = 1.4$, $p > .05$). The mean correctness scores (Table 2) for each stress type obtained by Chinese learners are significantly lower than those of the native Dutch group ($F [1, 38] = 39.4$, $p < .05$). This means that the native speakers of Dutch outperformed the Chinese speakers of Dutch. Patterns A and F are easier to identify, while D is the most difficult pattern.

Table 1: Mean correctness (Cor) and confidence (Conf) scores of Chinese (CSD) and Dutch subjects (NSD) broken down by accent type (A: word stress, B: compound, C: proper names, D: eventive sentences, E: non-eventive sentences, F: focus). For explanation of negative confidence scores see text.

Accent type	Language group	Mean		SD		N	Proficiency group	Mean		SD		N
		Cor	Conf	Cor	Conf			Cor	Conf	Cor	Conf	
A	CLD	3.95	1.45	1.60	2.17	20	CLD-L	3.20	.25	1.69	1.82	10
	NLD	5.80	4.38	.41	.50	20	CLD-H	4.70	2.65	1.16	1.84	10
B	CLD	4.10	1.50	1.25	1.81	20	CLD-L	3.70	.70	.95	.94	10
	NLD	5.15	3.38	.67	.96	20	CLD-H	4.50	2.30	1.43	2.14	10
C	CLD	3.90	1.38	1.29	1.79	20	CLD-L	3.10	.15	.99	1.03	10
	NLD	5.25	3.06	.79	1.04	20	CLD-H	4.70	2.62	1.06	1.52	10
D	CLD	3.55	.80	1.23	1.76	20	CLD-L	3.00	−.05	1.05	1.13	10
	NLD	4.75	2.73	.85	1.05	20	CLD-H	4.10	1.65	1.20	1.91	10
E	CLD	3.45	.71	1.23	1.74	20	CLD-L	2.90	−.18	1.29	1.53	10
	NLD	4.90	2.67	1.02	1.36	20	CLD-H	4.00	1.60	.94	1.51	10
F	CLD	4.45	1.98	1.05	1.59	20	CLD-L	4.20	1.32	1.03	1.17	10
	NLD	5.30	3.52	.80	1.11	20	CLD-H	4.70	2.63	1.06	1.73	10

Table 2: Overall mean correctness (Cor) and confidence (Conf) scores of Chinese and Dutch subjects

Groups	Mean		SD		N
	Cor	Conf	Cor	Conf	
Chinese	3.90	1.30	0.84	1.37	20
CLD-L	3.35	0.36	.55	.65	10
CLD-H	4.45	2.24	.72	1.27	10
Dutch	5.19	3.29	0.37	0.47	20

As Table 1 shows, native listeners of Dutch (NLD) obtained mean correctness scores above 5 points except for accent types D and E, whose mean scores are 4.8 and 4.9 respectively. The highest mean correctness score is 5.8 which goes to type A and the second highest (5.3) is type F. In the group of Chinese listeners of Dutch (CLD), the mean correctness scores are all below 5 points. The mean correctness scores of types B and F are comparatively higher (4.1 and 4.5, respectively), and the lowest is 3.5 which goes to type E. Both the CLD) and NLD groups have comparatively lower mean correctness scores for types D and E, and comparatively higher mean correctness scores for type F (the highest for CLD). The NLD group got the highest mean score for type A while this is not the highest for the CLD group. CLD got a comparatively higher mean score for Type B. This means that Chinese speakers of Dutch have problems with the identification of correct word stress. But their ability to judge the correctness stresses of compound words and focus information in sentences is higher than that of other types. Both groups have problems with the accentuation of enventive and non-eventive sentences.

We also conducted an RM-ANOVA using accent type as a within-subjects variable and different proficiency groups as between-subjects factor (for CLD listeners only). Correctness scores (Table 1) are significantly different across accent types ($F [5, 90] = 2.3$, $p = .05$). Pairwise comparisons reveal that accent types D and F differ significant ($p = .004$). There is no significant interaction between accent type and proficiency group ($F [5, 90] < 1$). The mean correctness scores (Table 2) for each stress type obtained by CLD-L are significantly lower than those of CLD-H ($F [1, 18] = 14.6$, $p < .05$). That means the CSD-H group outperformed the CLD-L group. Pattern F is easiest to identify, while Pattern D is difficult for Chinese listeners.

From Table 1, we also learn that CLD-H subjects got mean correctness scores above 4 points for all six accent types while CLD-L subjects obtained mean correctness scores around 3 points for types A, B, C, D and E while only accent type F got a significantly better score (4.2). Both CLD-L and CLD-H subjects got the highest mean correctness score for type F, and comparatively lower mean correctness scores for types D and E. Not surprisingly, CLD-L subjects have more problems with the correct identification of word stress. Both CLD-L and CLD-H have problems with the accentuation of eventive and non-eventive sentences (as do the native speakers of Dutch). All groups have significantly higher mean correctness scores for accent type F. Apparently, identifying the correct accentuation of sentences with focus information is relatively easy.

3.2. Analysis and discussion of confidence scores

For the analysis of confidence scores, each response was weighed positively if the accompanying judgment was correct. Otherwise, it was weighed negatively. We computed mean confidence scores per subject across accent patterns. We want to know how confident the subjects were when they made their judgments.

We again conducted an RM-ANOVA using accent types as within-subjects variables and different language groups as between-subjects factor. Confidence scores (Table 1) are significantly different across accent types ($F [5, 190] = 7.4$, $p < .05$). Pairwise comparisons reveal that the difference between accent types A and {C, D, E} are significant ($p < .05$), as is the difference between types {D, E} and F. There is no significant interaction between accent type and language group ($F [5, 190] =1.8$, $p > .05$). The mean confidence scores (Table 2) per stress type obtained by the Chinese subjects are significantly lower than those of the Dutch counterparts ($F [1, 38] = 37.4$, $p < .05$). This means that the native Dutch listeners were more confident than the Chinese listeners when they made their judgments. Both language groups were more confident when they judged types A and F and confident for types D and E.

As Table 1 shows, the mean confidence scores of Dutch group are all about 3 points except types D and E whose mean confidence scores are 2.7 and 2.7 (the lowest) respectively. NLD are most confident in their judgment of type A (4.4), followed by their judgment of type F (3.52). The mean confidence scores of CLD are above 1 point for all accent types except type D (0.8) and type E (0.7, the lowest).

An RM-ANOVA with accent type as a within-subjects variable and proficiency as a between-subjects factor show that confidence scores (Table 1) differ significantly across accent types ($F [5, 90] = 2.6$, $p < .05$). Only the difference between types F and D is significant (pairwise comparison, $p < .05$). There is no significant interaction between accent type and proficiency ($F [5, 90] < 1$). Mean confidence scores (Table 2) for each stress type obtained by the CLD-L are significantly lower than those of the CLD-H ($F [1, 18] = 17.4$, $p < .05$). This means that CLD-H were more confident in their judgments than CLD-L. Both language groups were more confident judging type F and less when judging type D.

Table 1 shows that CLD-H are most confident judging types A (2.7), C (2.6) and F (2.3), but less confident with types D (1.7) and E (1.6, lowest). Not surprisingly, CLD-L are not confident when judging type A (0.3). Like CLD-H subjects, they are not confident in their judgment of types D (−0.05) and E (−0.18).

Comparing mean correctness and mean confidence scores for the six accent types, we find that the two sets of scores correlate strongly. Nevertheless, for the Dutch group, the mean correctness score of type D is the lowest but the corresponding confidence score is not (the lowest confidence is for type E). This means though NLD got relatively higher mean correctness scores for type E than for type D, they are not sure whether their judgment was correct or not. Interestingly, for the Chinese listeners, whether CLD-L or CLD-H, the correctness scores and their confidence scores are in agreement with each other across all types of accentuation. It follows from the above comparisons, that all listeners took the experiment seriously and that their knowledge of Dutch accentuation patterns is truly reflected by their judgments.

4. CONCLUSIONS

Native speakers' correctness scores were significantly better than the scores by the high-proficiency Chinese speakers of Dutch, and the correctness scores of the latter were significantly higher than those of low proficiency Chinese learners of Dutch. This strongly suggests that greater proficiency in the language improves learners' ability to identify correct accent locations. Along with the correctness scores, confidence ratings decreased significantly from the native group to the high proficiency Chinese group and to the low proficiency Chinese group.

There were tantalizing differences between the linguistic categories, showing that some accent patterns are easier to judge − and thus learn − than others. Native speakers find it easiest to identify primary word stress (5.8 correct, where 6.0 is the highest possible score), but Chinese L2 speakers of Dutch find it easiest to judge the correctness of the focus condition (4.5). This reflects the fact that the location of primary and secondary word stress in the Dutch words we used is a language specific, arbitrary fact, whereas the focus structure of the sentences we used is given by the context. The Chinese learners thus derived the correct accentuation pattern in the focus condition from the pragmatics of the mini-dialogues that constituted the stimuli. Judging accentuation of eventive and non-eventive sentences is the most difficult for both native speakers (4.8) and for Chinese L2 speakers of Dutch (3.5). This is understandable, since both correct and incorrect accentuations in fact constitute natural accentuation patterns, given some adjustments in the context. Overall, with the exception of accentuation as a function of focus, our results show that the accentuation patterns of Dutch sentences are language specific and must be learnt.

ACKNOWLEDGEMENTS

This research was financially supported by the Netherlands Organisation for Research (NWO) as part of the programme 'Intonation in varieties of Dutch' (principal investigator Gussenhoven). The authors thank Judith Hanssen for permitting them to record her as the second speaker of our stimulus materials.

REFERENCES

Chen Hua 2008. *On Chinese EFL learner's English intonation patterns*. Shanghai: Shanghai Foreign Language Education Press.

Ellis, R. 1994. *The study of second language acquisition*. Oxford: Oxford University Press, 482-492

Gao Qiaoyang 1984. Grammatical Stress in Mandarin. *Language Teaching and Research*, 2, 65-76.

Booij, G. 1999. *The phonology of Dutch*. Oxford: Oxford University Press.

Goldsmith, J. 1976. Autosegmental tonology. Ph.D. Dissertation, M.I.T.

Gussenhoven, C. 1999. Illustrations of the IPA: Dutch. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press. 74-77.

Gussenhoven, C. 1983. Focus, mode and the nucleus. *Journal of Linguistics* 19, 377-417.

Gussenhoven, C. (ms) Sentential prominence in English. To appear in Oostendorp, M.; Ewen, C.; Hume, B., Rice, K. (eds.), *Companion to Phonology*. Wiley-Blackwell.

Hyman, L.M. 1978. *Tone and/or accent*. In Napoli, D.J. (ed.), *Elements of tone, stress and intonation*. Washington: Georgetown University Press. 1-20.

Gut, U., Trouvain, J., Barry, W.J. 2007. Bridging research on phonetic descriptions with knowledge from teaching practice: The case of prosody in non-native speech. In Trouvain, J., Gut, U. (eds.), *Non-native prosody, phonetic description and teaching practice*. Berlin: Mouton de Gruyter. 3-25.

Ladd, D.R. 2008. *Intonational phonology*. Cambridge: Cambridge University Press.

Schmerling, S. 1976. *Aspects of English sentence stress*. Austin: University of Texas Press.