THE RELATION BETWEEN PITCH ACCENT TYPES, HEAD MOVEMENTS AND PERCEIVED PROSODIC PROMINENCE IN L2 FRENCH

Florence Bailly[1,2], Stefan Baumann[1], Patrick Rohrer[2]

[1]IfL Phonetik, University of Cologne
[2] GreP-G, Universitat Pompeu Fabra
{fbailly; stefan.baumann}@uni-koeln.de, patrick.rohrer@upf.edu

ABSTRACT

Recent research has shown the relevance of multimodal cues in the realization of prominence in discourse. Speakers may not only use prosodic cues - e.g. pitch accents - to stress important information but also visual cues - e.g., manual and non-manual gestures such as head movements - in synchrony.

As part of a larger project comparing the multimodal marking of information structure by L1 and L2 speakers, this study reports on the relationship between perceived prominence in L2 speech and the pitch accent and head movement types used by 25 Catalan learners of French during a narrative task. Results confirm the relationship between pitch accents, gestural cues, and prosodic prominence in L2 learners and show that higher prosodic prominence is associated with rising, falling, and high pitch accents as well as with protrusions and nods of the head. Falling contours associated with highly prominent words were less marked by head movements, indicating potential differences from L1 speech.

Keywords: prominence, prosody, head movements, pitch accents, L2 speech

1. INTRODUCTION

Prominence refers to the level of perceptual strength of a constituent in relation to other constituents in the same utterance, based on its structural and acoustic properties. In Autosegmental-Metrical phonology [1], prominence is associated with metrically strong positions in an utterance, with the rightmost position, called “nuclear”, being the strongest. “Prenuclear” prominences occur earlier in the phrase (often in initial position) and are thus regarded to be structurally weaker. Various prominence rating studies on American English and German confirmed this broad picture since they found a generally higher level of *perceived* prominence for nuclear accents in comparison with prenuclear accents and unaccented words. As to the *type* of accent in nuclear position, both languages show higher prominence values for rising accents, followed by high, falling and low pitch accents [2,3].

Much like West Germanic languages, Catalan and Spanish are stress-accented languages: the syllables bearing primary stress usually serve as the landing site for pitch accents, which depend on the larger prosodic structure [4]. Actual prosodic prominence surfaces on primarily stressed syllables at the word level and accents at the phrase level, with F0 movements being an essential cue of the latter [5]. Deaccentuation is rare. In French, by contrast, there is no word stress. Accent placement is believed to be constrained by prosodic phrasing and to signal phrase structure, with a compulsory high or low phrase-final accent and an optional initial rise [6].

There is evidence on the temporal association between prosodic prominence (pitch-accented syllables) and the prominence (strokes or apexes) of referential and non-referential hand gestures [7, 8, 9, 10, 11]. Regarding non-manual gestures, studies have found that head gestures tend to co-occur with pitch-accented syllables [12, 13, 14, 15]. Crucially, Alexanderson et al. [13] found that head nods co-occurred with words bearing a focal accent in Swedish spontaneous dialogic speech. In parallel, there is also direct evidence of the relationship between perceived auditory prominence and non-manual gestures. Swerts and Krahmer [16] investigated both head and eyebrow movements in a Dutch news reading corpus: Strongly prominent words tended to co-occur with both head and eyebrow movements (67%), while weak prominence was mostly produced without gesture (47%), or with only-head (16%) or only-eyebrow movement (3%).

Little is known about the use of multimodal cues and their relationship with perceived prominence in L2 speech. McCafferty [17] suggested that L2 learners with lower proficiency may use non-referential (beat) gestures to parse words into syllables and control their fluency. Importantly, it is likely that lower proficiency L2 speakers might be using gestures instead of other linguistic devices for a range of pragmatic functions including the marking of salient information [18], in a similar way as children may do [19]. Our goal is to explore for the first time the relationship between perceived prominence and prosodic and gestural cues in L2 speakers by answering the following research questions:

- Which types of pitch accent are perceived as more prominent in L2 speech? We expect high and rising pitch accents to be perceived as more prominent than low and falling pitch accents.
- Which types of head movement are accompanied by higher prosodic prominence? To date, only nods have been specifically studied in relation to prominence. Here we investigate other types of head movement in an exploratory part of the study.
- What is the relationship between pitch accent types and head movement types, and is this relationship reflected in their association with perceived prominence?

2. METHODS

2.1. Participants and elicitation task

Participants were 25 bilingual speakers of Catalan and Spanish (M_{age} = 19.6 years, *SD* = 0.91) learning French. They were all second-year undergraduate students in Translation and Interpretation Studies. French was their second foreign language after English. Their self-reported proficiency in French varied between A2 (20%), B1 (44%) and B2 (36%).

During the semester, they had to complete two assignments as a means of data collection for this study. They were asked to video-record themselves producing a short spontaneous monologue 1) about their Erasmus stay, which took place the previous year. In order to obtain comparable oral productions, the instructions entailed a series of questions, as follows: *"Please explain your Erasmus stay in a few sentences in French. Where did you go and when did you leave? Did everything go smoothly? What are your best memories? Would you like to live in a foreign country again?"* and 2) describing their best friend, with the following instructions: *"Explain in a few sentences in French who your best friend is. How is he/she physically? How did you meet? What do you like and dislike about him/her?"*

For the two assignments, participants had to follow a link sent by the experimenter which led them to an online survey platform where they could read the instructions for video recording and for the task. No indication was given regarding head or body movement. They were asked to use their laptop camera and the online recording software https://webcamera.io/ (quality 720p) and then share the file with the teacher and the experimenter.

2.2. Annotations

Perceived prominence and prosodic annotations. The auditory information of the 50 video recordings produced by the participants was extracted as sound files (WAV 48KHz 16-bit) using Adobe Premiere Pro. The sound files were annotated automatically in Praat [20] for word, syllable and phoneme with EasyAlign [21] and manually corrected by the first author. In a first step, prosodic and prominence annotations were carried out independently of each other. Prominence was annotated auditorily following the guidelines proposed by the DIMA annotation system [22], that is, weak, strong or extra strong prominence labels (levels 1 to 3) were attributed to words according to their perceived salience in the utterance. Then, phrasing and pitch accents were annotated. As L2 speech presents deviations from the prosodic patterns that are predicted by F_ToBI phonological annotations [23], it was decided to follow a more phonetic annotation of the pitch accents in the accentual phrases (AP) when deemed necessary, as advocated by Hualde & Prieto [24]. Each AP contained one obligatory pitch accent at the right edge of the phrase (L*, HL*, !H*, H*L, H*, L*H, or LH*) and optionally a high or rising initial accent (Hi). Inter-annotator reliability between the first two authors (one of them being a native French speaker) on 20% of the data was calculated for prominence ratings. A moderate score of .475 Cohen's Kappa was obtained.

Head movements. The video recordings were coded in ELAN [25] for non-referential head movements only, i.e., we excluded head movements conveying any semantic meaning such as shaking the head rapidly from left or right to express negation, or up and down repeated movements for approval. However, when a head movement had a semantic and a rhythmic component, it was decided that it would be annotated and included in the analysis, for example a protrusive movement of the chin to one side to indicate a specific location. We applied the M3D annotation scheme, which specifies five head movement types [26; 27, p. 25]: *Nod*, up and down movements; *Turn* right and left movements; *Tilt*, the top of the head goes in one direction (left or right) and the chin goes in the opposite direction; *Slide*, displacement of the whole head to the left or right; *Protrusion*: displacement of the whole head forward or backward.

The gesture strokes were detected and annotated by searching for the beginning and end of the most salient movement of the head in a frame-by-frame observation and without the audio, in order to avoid any influence from the speech stream. The type of head movement was chosen after observation of the movement at normal speed. Ten percent of the data was annotated by the first two authors and inter-annotator reliability was assessed for gesture stroke and type labelling with the built-in inter-annotator

reliability tool in ELAN, which uses an algorithm to assess both temporal overlap and the values assigned to the annotations. The algorithm returned kappa values of .65 for the identification of each type of phase, indicating substantial reliability.

Finally, the prominence and prosodic annotations in Praat were imported into ELAN which allows the creation of a time-aligned database for further processing in R [28]. Phrase-final pitch accents were grouped according to their shape: low (L*), falling (H*L, HL*), high (H*), and rising (L*H, LH*, HH*) - plus the phrase-initial high or rising accent (Hi).

2.3. Statistical analysis

Three Generalized Linear Models (GLM1, GLM2, and GLM3) were run in R with the *lme4* package [29]. A Poisson regression was used to model count data and to account for between-category differences (the unbalanced number of occurrences), the models were offset by the total number of items for each category. GLM1 assessed the association between perceived prominence and pitch accent type. The dependent variable was the number of syllables annotated with a pitch accent and the fixed factors were PROMINENCE (2 levels: weak, strong; 'extra strong' did not occur in the data) and PITCH ACCENT TYPE (5 levels: initial accent, low, falling, high, rising) as well as their interaction. GLM2 evaluated the association between types of movement and perceived prominence. The dependent variable was the number of syllables annotated which contained the apex of the head movement and the fixed factors were PROMINENCE (2 levels: weak, strong) and HEAD MOVEMENT TYPE (5 levels: nod, protrusion, slide, tilt, turn) as well as their interaction. GLM3 assessed the association between types of pitch accent and types of head movement, in GM3, the dependent variable was the number of syllables annotated with both a head movement and a pitch accent, and the fixed factors were PITCH ACCENT TYPE (5 levels: initial accent, low, falling, high, rising) and HEAD MOVEMENT TYPE (5 levels: nod, protrusion, slide, tilt, turn) as well as their interaction.

3. RESULTS

Perceived prominence and pitch accent types. Results of GLM1 show a significant effect of PITCH ACCENT TYPE ($\chi^2(4) = 1002.49$, $p < .001$ and an interaction PROMINENCE X PITCH ACCENT TYPE ($\chi^2(4) = 263.35$, $p < .001$). Post-hoc comparisons show that falling and rising pitch accents are related to significantly more prominent rating scores followed by high, low, and initial accents (see Fig.1).

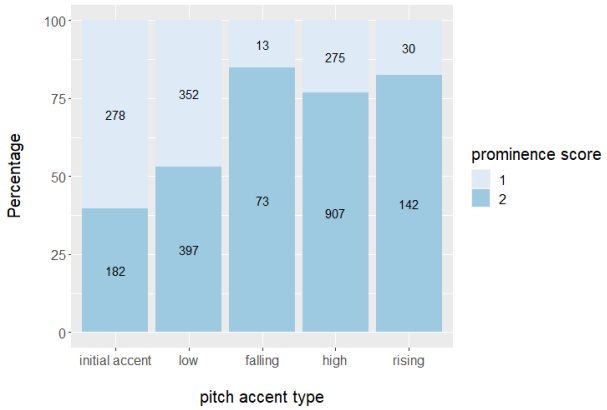

Figure 1: Prosodic prominence relating to the type of pitch accent (percentage along y-axis; N of occurrences indicated in graph).

Perceived prominence and head movement types. Results of GLM2 show a significant effect of HEAD MOVEMENT TYPE ($\chi^2(4) = 40.05$, $p < .001$) and an interaction PROMINENCE X HEAD MOVEMENT TYPE ($\chi^2(8) = 35.83$, $p < .001$). Post-hoc results indicate that protrusions, nods, and slides are associated with high (level 2) prosodic prominence (see Fig. 2). Across types of movements, protrusions are significantly associated with higher prominence compared to tilts and slides. Other differences are not significant.

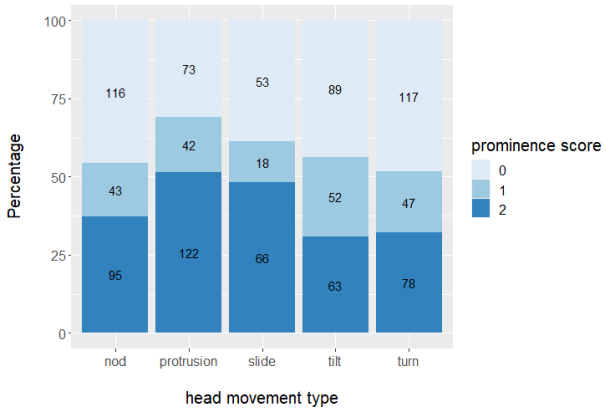

Figure 2: Prosodic prominence levels co-occurring with gesture types (percentage along y-axis; N of occurrences indicated in graph).

Types of pitch accent and types of head movement. Results of GLM3 show a significant effect of PITCH ACCENT TYPE ($\chi^2(4) = 649.09$, $p < .001$), HEAD MOVEMENT TYPE ($\chi^2(4) = 45.31$, $p < .001$) and an interaction PITCH ACCENT TYPE X HEAD MOVEMENT TYPE ($\chi^2(16) = 30.12$, $p < .05$). Post-hoc results reveal that all types of head movement co-occur the most with high pitch accents followed by low pitch accents. Across categories protrusions are more likely to co-occur with a high accent compared to tilts and slides, but no other difference is significant.

In terms of proportions (see Table 1), protrusions and slides are the most frequent head movements co-

occurring with high pitch accents (around 59%). Low accents are more frequently accompanied by tilts and turns (around 30-32%), which are themselves co-occurring with lower levels of prosodic prominence. Altogether, rising accents, even if less frequent, are the ones that are more often co-occurring with a head movement (59%, most of which were nods) while falling accents, despite being also perceived as more prominent, tend to be less frequently accompanied by a head movement (around 26%). Phrase-initial accents were least often co-occurring with a head movement (15%).

4. DISCUSSION AND CONCLUSION

In this study, we asked three questions related to the expression of prominence through prosodic and gestural cues in Catalan L2 learners of French. The first analysis shows that falling, rising and high phrase-final accents in L2 French were perceived as more prominent than phrase-final low and phrase-initial accents. Falling and rising accents are significantly more prominent than high accents, which might indicate that pitch *movement* was decisive for the annotators' prominence judgments.

Regarding gestural cues, the second analysis showed that protrusions, nods, and slides coincide the most with higher prosodic prominence. Around 70% of all protrusions are accompanied by prosodic prominence level 1 (weak) or 2 (strong), with about 50% of level 2 prominences. Future research should investigate whether the prominence of the head movement depends more on the amplitude and duration of the movement, as protrusions and nods may allow for a more expanded movement.

The third analysis shows that when a head movement co-occurs with a pitch accent, it is predominantly with high accents, regardless of the head movement type. It seems that particular contours may attract more or less head movements. While falling and rising contours were perceived both as more prominent, the former attracted only half the number of head movements compared to the latter. On the one hand, this contradicts the cumulative cue hypothesis [30] claiming that the relation between prosodic and gestural cues to prominence should be additive in nature, on the other it suggests that L2 learners may not yet be completely able to produce multimodal prominence and have difficulties at least with certain types of contours. The lack of association between initial accents and head movement may indicate that phrase-final accents generally attract more head movements as prominence markers in our speakers. Future studies will include L1 controls to assess these issues.

This study provides evidence that the head movements which are associated with higher prosodic prominence (protrusions, nods and slides) more frequently accompany pitch accents reflecting higher prosodic prominence, particularly rising and high pitch accents. These results are in line with [31], who found that protrusions and nods tend to co-occur with new(er) information in discourse [32], which is usually signalled by prosodic prominence. Crucially, a vast majority of the total number of head movements co-occurs with pitch accents, showing the close association between prosody and gesture (Table 1). To continue assessing the role of multimodal cues in prominence, our results need to be contrasted with a perceptual evaluation of the prominence of head movements alone and of both visual and acoustic cues together, including their relations with information structure marking.

By exploring the relationship between multimodal cues and prominence, we can gain further insight on the developmental path of speech production in an L2. For instance, French speakers may use gestural cues more to indicate discourse structure [33]. This study constitutes the first step of a larger study comparing Catalan learners of French and native French speakers to assess these differences.

	protrusions	nods	turns	Tilts	slides	total	corpus
high	126 (59,4%)	98 (48%)	94 (48,7%)	67 (47,9%)	68 (59,6%)	453 (38,3%)	1182
low	40 (18,9%)	50 (24,5%)	59 (30,6%)	45 (32,1%)	23 (20,2%)	217 (29 %)	749
rising	23 (10,8%)	32 (15,7%)	21 (10,9%)	13 (9,3%)	13 (11,4%)	102 (59,3%)	172
init. acc.	16 (7,5%)	19 (9,3%)	18 (9,3%)	12 (8,6%)	4 (3,5%)	69 (15%)	460
falling	7 (3,3%)	5 (2,4%)	1 (0,5%)	3 (2,1%)	6 (5,3%)	22 (25,6%)	86
total	212 (86,1%)	204 (80,3%)	193 (79,7%)	140 (68,6%)	114 (83,2%)		
corpus	237	254	242	204	137		

Table 1: Number of occurrences and proportions (in %) of pitch accent types and head movement occurring jointly and in relation to the total number of pitch accent and head movement types in the corpus. Note: init. acc = initial accent.

5. REFERENCES

[1] Ladd, R. 2008. *Intonational phonology*. Cambridge University Press.

[2] Baumann, S., Röhr, C. 2015. The perceptual prominence of pitch accent types in German. *Proc. 18th ICPhS Glasgow*, paper 298.

[3] Cole, J., Hualde, J. I., Smith, C., Eager, C., Mahrt, T., de Souza, R. N. 2019. Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *J. Phonetics 75*, 113–147.

[4] Prieto, P. 2014. The intonation phonology of Catalan. In S.-A. Jun (ed). *Prosodic typology II.* Oxford University Press.

[5] Ortega-Llebaria, M., Prieto P. 2007. Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In P. Prieto, J. Mascaró, & M.-J. Solé (eds). *Segmental and prosodic issues in Romance phonology.* John Benjamins, 155–176.

[6] Jun, S.-A., Fougeron, C. 2000. A phonological model of French intonation. In A. Botinis (ed), *Intonation: Analysis, modeling and technology.* Springer, 209–242.

[7] Esteve-Gibert, N., Prieto, P. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *J. Speech Lang. Hearing Res. 56*, 850–864.

[8] Leonard, T., Cummins, F. 2011. The temporal relation between beat gestures and speech. *Lang Cog. Proc. 26*, 1457–1471.

[9] Yasinnik, Y., Renwick, M., Shattuck-Hufnagel, S. 2004. The timing of speech-accompanying gestures with respect to prosody. *J. Acoust. Soc. Am. 115*, 2397.

[10] Loehr, D. P. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Lab. Phon. 3*, 71–89.

[11] Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Front. Psychology 9*, Article 1514.

[12] Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swets, M., Prieto, P. 2017. The timing of head movements: The role of prosodic heads and edges. *J. Acoust. Soc. Am. 141*, 4727–4739.

[13] Alexanderson, S., House, D., Beskow, J. 2013. Aspects of cooccurring syllables and head nods in spontaneous dialogue. *Proc. 12th AVSP conference Annecy*, 169–172.

[14] Ambrazaitis, G., House, D. 2017. Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Comm. 95*, 100–113.

[15] Liu, Y., Shamei, A., Chow, U.Y., Soo, R., Pineda Mora, G., de Boer, G., Gick, B. 2020. F0-related head movement in blind versus sighted speakers. *J. Acoust. Soc. Am. 148*, 190–194.

[16] Swerts, M., Krahmer, E.J. 2010. Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions. *J. Phon. 38*, 197–206.

[17] McCafferty, S. G. 2006. Gesture and the materialization of second language prosody. *IRAL 44*, 197–209.

[18] Graziano, M., Nicoladis, E., Merentette, P. 2020. How referential gestures align with speech: Evidence from monolingual and bilingual speakers. *Lang. Learn. 70*, 266–304.

[19] Esteve-Gibert, N., Loevenbruck, H., Dohen, M., and D'Imperio, M. 2021. Preschoolers use head gestures rather than prosodic cues to highlight important information in speech. *Dev. Sci. 25*: e13154.

[20] Boersma, P., & Weenink, D. 2022. Praat: doing phonetics by computer [Computer program]. Version 6.2.06, retrieved 15 March 2022 from https://www.praat.org

[21] Goldman, J.-P. 2011. EasyAlign: an automatic phonetic alignment tool under Praat. *Proc. InterSpeech 2011 Florence*. ISCA Archive., 3233–3236.

[22] Kügler, F., Baumann, S., Röhr, C.T. 2022. Deutsche Intonation, Modellierung und Annotation (DIMA) – Richtlinien zur prosodischen Annotation des Deutschen. In C. Schwarze & S. Grawunder (eds), *Transkription und Annotation gesprochener Sprache und multimodaler Interaktion*. Narr. 23–54.

[23] Delais-Roussarie et al. 2015. Intonational phonology of French: Developing a ToBI system for French. In S. Frota & P. Prieto (eds), *Intonation in Romance.* Oxford University Press. 63–100.

[24] Hualde, J., Prieto, P. 2016. Towards an International Prosodic Alphabet, *Lab. Phon. 7*, 5.

[25] ELAN Version 6.4 [Computer software]. 2022. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive.

[26] Wagner, P. 2014. Gesture and speech in interaction: An overview. *Speech Comm. 57*, 209–232.

[27] Rohrer, P., Vilà-Giménez, I., Florit-Pons, J., Esteve-Gibert, N., Ren, A., Shattuck-Hufnagel, S., Prieto, P. 2020. *The MultiModal MultiDimensional (M3D) labeling system.* https://doi.org/10.17605/OSF.IO/ANKDX

[28] R Core team. 2022. R Version 4.2.1. https://www.R-project.org/.

[29] Bates. D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat, Soft. 67*, 1–48.

[30] Ambrazaitis, G., House, D. 2022. Probing effects of lexical prosody on speech-gesture integration in prominence production by Swedish news presenters. *Lab. Phon. 13*, 1–35.

[31] Baills, F. & Baumann, S., Submitted. Multimodal marking of information structure in the L2: the case of Catalan learners of French.

[32] Riester, A. & Baumann, S. 2017. The RefLex Scheme – Annotation Guidelines. SinSpeC. Working Papers of the SFB 732, vol. 14. University of Stuttgart.

[33] Rohrer, P. 2022. *A temporal and pragmatic analysis of gesture-speech association* [PhD thesis, UPF]. TDX.