

Prosody and interactional fluency of Italian learners of German

Simona Sbranna

Eurosla Studies 8



EuroSLA Studies

Editor: Amanda Edmonds, Université Côte d'Azur

Associate editors:

Gabriele Pallotti, University of Modena and Reggio Emilia

Ineke Vedder, University of Amsterdam

In this series:

1. Pérez Vidal, Carmen, Sonia López-Serrano, Jennifer Ament & Dakota J. Thomas-Wilhelm (eds.). Learning context effects: Study abroad, formal instruction and international immersion classrooms
2. Saturno, Jacopo. Utterance structure in initial L2 acquisition.
3. Bardel, Camilla & Laura Sánchez (eds.). Third language acquisition: Age, proficiency and multilingualism.
4. Edmonds, Amanda, Pascale Leclercq & Aarnes Gudmestad (eds.), Interpreting language-learning data.
5. Berthele, Raphael & Isabelle Udry (eds.). Individual differences in early instructed language learning: The role of language aptitude, cognition, and motivation.
6. McManus, Kevin & Monika S. Schmid (eds.). How special are early birds? Foreign language teaching and learning.
7. Granget, Cyrille, Isabel Repiso & Guillaume Fon Sing (eds.). Language, creoles, varieties: From emergence to transmission.
8. Sbranna, Simona. Prosody and interactional fluency of Italian learners of German.

Prosody and interactional fluency of Italian learners of German

Simona Sbranna

Simona Sbranna. 2025. *Prosody and interactional fluency of Italian learners of German* (Eurosla Studies 8). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/515>

© 2025, Simona Sbranna

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/> 

The work presented in this book is based on the author's doctoral dissertation, which was accepted by the Faculty of Arts and Humanities at the University of Cologne in 2023.

ISBN: 978-3-96110-530-4 (Digital)

978-3-98554-150-8 (Hardcover)

ISSN: 2626-2665

DOI: 10.5281/zenodo.15726570

Source code available from www.github.com/langsci/515

Errata: paperhive.org/documents/remote?type=langsci&id=515

Cover and concept of design: Ulrike Harbort

Typesetting: Simona Sbranna, Hannah Schleupner, Sebastian Nordhoff

Proofreading: Bruno Behling, Sebastian Nordhoff

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: Xe_{La}T_EX

Language Science Press

Scharnweberstraße 10

10247 Berlin, Germany

<http://langsci-press.org>

support@langsci-press.org

Storage and cataloguing done by FU Berlin

Contents

Acknowledgments	v
1 Introduction	1
1.1 Communicative competence in oral performance	2
1.1.1 The theory relating to prosodic competence	4
1.1.2 The theory relating to interactional competence	5
1.1.3 The gap between the theory and practice of language teaching and learning	8
1.2 Outline of the book	9
1.3 Corpora	11
1.3.1 Participants	11
1.3.2 Data collection	13
1.3.3 Data use across the studies	15
2 Prosodic marking of information status	17
2.1 Background	18
2.1.1 Information structure	18
2.1.2 Prosodic encoding of information structure	19
2.1.3 Categorical and continuous approaches	20
2.1.4 Prosodic marking of information structure in Italian . .	23
2.1.5 Prosodic marking of information structure in German .	24
2.1.6 Information structure and prosody in L2	25
2.1.7 A comparative research programme on Italian and German as L1 and L2	27
2.2 Method	30
2.2.1 Data elicitation	30
2.2.2 Corpus	33
2.2.3 Measurements	37
2.2.4 Bayesian analysis	39
2.3 Results: Italian L1	40
2.3.1 Learners of German	40
2.3.2 Monolingual control group	43

Contents

2.3.3	Summary	44
2.4	Perceptual validation for Italian L1	46
2.4.1	Procedure	46
2.4.2	Results	48
2.4.3	Acoustic analysis of correctly identified items	48
2.4.4	Discussion	52
2.4.5	Summary	54
2.5	Results: German L1	54
2.5.1	Summary	60
2.6	Results: German L2	60
2.6.1	Proficiency levels	63
2.6.2	Summary	66
2.7	Interlanguage compared to learners' native and target languages	67
2.7.1	F0 contours	68
2.7.2	Mass	70
2.7.3	Discussion	71
2.7.4	Summary	73
2.8	Implications for phonological analysis	73
2.8.1	Background	74
2.8.2	Alignment and scaling	75
2.8.3	Duration and Intensity	78
2.8.4	A phonological interpretation	82
2.8.5	Summary and discussion	87
2.9	Conclusion	89
3	Turn-taking fluency in L2 interactions	93
3.1	Background	93
3.1.1	Assessing interactional competence	93
3.1.2	Fluency in L2 studies	95
3.1.3	Fluency measures and operationalisations	97
3.1.4	Interactional fluency visualisation methods	101
3.2	Method	103
3.2.1	Corpus	103
3.2.2	Elicitation method	104
3.2.3	Metrics	106
3.2.4	Procedure for tools generation	107
3.2.5	Bayesian analysis	111
3.3	Quantitative analysis	113

3.4	Qualitative analysis	115
3.4.1	Procedure	115
3.4.2	Results	117
3.5	Cross-linguistic comparison	119
3.6	Conclusion	120
4	Backchannels in L2 interactions	125
4.1	Background	125
4.1.1	Backchannel contribution to interactional fluency	125
4.1.2	Backchannel definitions and categorisations	127
4.1.3	Backchannel use across languages and cultures	129
4.1.4	Backchannels in L2 speech	130
4.2	Method	133
4.2.1	Corpus	133
4.2.2	Procedure and Metrics	134
4.3	Results	136
4.3.1	BC frequency	136
4.3.2	BC length	138
4.3.3	BC structure	140
4.3.4	BC type	141
4.3.5	BC type by function	144
4.3.6	Other VSU type by function	146
4.3.7	BC intonation	147
4.3.8	Other VSU intonation	150
4.4	Conclusion	151
5	Conclusion	155
5.1	Summary	156
5.1.1	Prosodic marking of information status	156
5.1.2	Turn-taking	157
5.1.3	Backchannels	158
5.2	Implications for second language research and teaching	159
5.3	Limitations and future directions	161
	References	183
	Index	213
	Name index	213

Acknowledgments

It is finished. The time has come to thank the people who made the writing of this book possible and who have accompanied me on this journey.

First of all, I would like to thank my supervisors, Martine and Elina, for the time they took to plan, discuss, and improve the content of this book, and for the knowledge and skills they passed on to me. Your contribution to my professional development has been immense, but I have also learned much more that will be useful to me in life. Thank you for your roles as three-hundred-and-sixty-degree trainers.

Thanks to another supervisor, Renata, who is no longer with us, but without whom I would not be here writing these acknowledgements. Thank you for inspiring me and setting me on this path. Thank you for being my mentor with the love of a mother. I will always carry you with me, and this book is also for you. I hope you can be proud of that little girl you encouraged to fly out of the nest.

Thanks to my older colleagues, Francesco and Simon, who are in my eyes like older brothers (yes, we Italians always get attached to everyone like family members). Francesco, back in 2015, when a lost little girl was looking for advice for her B.A. thesis during her semester abroad and came across a researcher in Germany who went to work in a suit, was Neapolitan, and knew her thesis supervisor. Practically unreal, like a mirage in the desert. And Simon, in his course on second language acquisition back in 2016. Thanks to both of you for the long discussions, not only on professional topics, from which part of the thesis originated, but above all for the moments of connection between human beings, which are the ones most fondly etched in memory. Also, for Simon, who was also the language editor of this text, a special thanks for his patience and care devoted to this book.

Thanks also to Aviad and Christine for taking the time, over and over again, to give me advice and offer explanations. Your input was crucial in the development of certain aspects of this book, and I am immensely grateful for that. To Christine in particular, a special thanks for all the lunch breaks shared during the pandemic in the almost-empty laboratory. Thank you, too, for the discussions about life, they are always the best.

Acknowledgments

A special thank you to Caterina, who has unwittingly been a big sister. I could write a page just about you and still not do justice to your contribution to the development of this thesis and my growth. I have learned so much from you, and once again, the most important lessons have been about life. Thank you, because in you I have also found a friend and confidante.

I would like to thank Anna Clelia, Celine, and Harriet, student assistants and great companions, for their affection towards me. Thanks to all the other lab members, Stefan, Theo, Tabea, Lena, Alicia, Maria, Malin, and the guest researcher Vincenzo (another great friend) for the mutual help, meals, and shared chats. Your positive energy contributed to a pleasant and genuine working environment.

Thanks to a distant colleague and one of my dearest friends, Loredana. Thank you for your moral support, your advice, and for affectionately welcoming my outbursts in the most stressful moments. Having someone who shares the same journey and can fully understand has been an immense help. Thank you for being there and, whether close or distant, I hope we will still go a long way together. Thank you to all my lifelong Italian friends, who have stayed by my side even if they do not see me very often: Pio, Susi, Daria, Francesca, Alfonso, Carolina, Laura, Sabrina, Grazia, Ilaria, Anna and Rossella. And thanks to the “new” friends (although for some it has already been eight years) who have made me feel at home in a foreign country with their affection and sincerity: Andre, Karen, Nele, Sivan and Giulia.

However, the greatest thank you goes to my family: my mum Santa, my dad Mimmo and my sister Noemi. Your support, despite the pain of being separated, has been an act of love and courage. Thank you for always encouraging and supporting me in my choices, even when you did not share them or suffered because of them. Thank you for passing on to me the values that allow me to act every day in the world, and for having taught me to always do my best and be the best version of myself. Thank you for your immense love; I feel it from kilometres away. I am truly blessed to have you as family, confidants, and best friends. I could not have asked for better. I love you all so much.

How could I forget to thank my German family, Gerhard? Thank you for accompanying me during these eight years and contributing so much to my thinking and philosophy of life. I have learned so much from you. Thank you for being my family here. When I felt lonely, I knew you were there and that I could count on you. That is an invaluable feeling for someone who lives far from their homeland. I hope I can give back at least some of the affection and selfless support I have received. I love you very much. Thank you.

And then there is you. This book also led me to you: my greatest and most unexpected gift. The man who cautiously knocked on the door of my heart and patiently waited for me to see him clearly. And when I did, I knew immediately that it was you. You were the one I was looking for. The man who awakened the joyful child in me. The man who knew how to make me fall in love with his gentle kindness. The man I want to share my path with. Thank you, my love, for making me feel safe and at home every day. Thank you for your love of few words but tender actions. Thank you for your gentleness and sensitivity. I love you so much. I thank this book for having brought us together.

Finally, I must thank one last person, Simona, myself, whom I often forget to thank. Thank you for your stubbornness, your strength, and your ability to always rise from the ashes. Know that I love you and am proud of you, whatever you do and wherever you go.

1 Introduction

This book is concerned with prosody and fluency phenomena of interactional competence (IC) in second language acquisition (SLA).¹

The particular skills investigated are a) prosodic marking of information status, b) floor management via turn-taking and c) backchanneling (i.e. vocal feedback signals), all in dyadic interactions. These abilities, which might seem unrelated, are actually complementary to each other. While the study on prosody will be concerned with learners' ability to master melodic aspects of the interaction, i.e. the intonational meaning (Wennerstrom 2006), the study of turn-taking and backchannels with a turn-regulating function will be focused on learners' ability to manage the rhythm of the interaction and ensure a fluent and smooth conversational tempo. For these reasons, prosody and fluency phenomena have often been investigated together (Trofimovich & Baker 2007; Tseng et al. 2005; Kallio et al. 2023; Ashby 2016). In this book, however, I approach fluency aspects with a novel perspective, that is, within the specific context of the coordinated collaborative interaction, as opposed to the tradition of studies which has regarded fluency from an individual perspective (Tavakoli & Skehan 2005; Segalowitz 2010; Kormos 2006). In line with this approach, the abilities of turn-taking and backchanneling contributing to the fluency of conversation will be grouped together under the broader definition of interactional competence (IC, Young 2014).

¹The terms *learning* and *acquisition* have traditionally been used with different meanings. The former has been used to describe a conscious process occurring in formal settings under formal instruction (i.e. L2 classrooms), while the latter has been used to describe a subconscious or unconscious process of assimilation occurring in naturalistic settings (i.e. language immersion through direct exposure, such as in the country where the L2 is the dominant language used for daily communication). However, recent research does not tend to differentiate between acquisition and learning, and the term acquisition, which I also adopt, is the most widespread. This is rooted in the "language socialisation paradigm" (Watson-Gegeo 2004) refusing the dichotomous distinction of the two processes (Rivers 1994; Ellis 1989). The classroom is seen as an inherent part of social life in many cultures and, therefore, should be considered as an integral part of a naturalistic setting (Watson-Gegeo & Nielsen 2003). This claim is supported by research findings indicating that the outcome of the two settings can be highly similar (Rivers 1994; Willett 1995), with important implications for language teaching, which should aim for more naturalistic methods and materials.

1 Introduction

Despite being crucially involved in oral communication, prosodic and interactional competence still receive limited attention, both within L2 research (for prosody: Derwing & Munro 2015; Gut 2009; Mennen 2004; for IC: Cekaite 2007; Galaczi 2014; Yamamoto et al. 2015) and the applied field of language teaching (for prosody: Aronsson 2014; Piccardo & North 2017; for IC: Campbell-Larsen 2022; Cohen 2005; Van Compernelle & Soria 2020). In fact, the poor mastery of prosodic and interactional competence has been reported to negatively impact learners' communicative abilities and perception. On the communicative side, a poor control of these abilities might slow down the processing of the information and reduce communication efficacy (Sørensen et al. 2019; Song & Iverson 2018), or even cause communication breakdowns (Sbranna et al. 2020). On the perceptual side, studies report a negative impact on the perception of learners' proficiency (Van Os et al. 2020; Trofimovich & Baker 2007), comprehensibility and intelligibility (Munro & Derwing 1999; Kang 2010; Hahn 2004) by native speakers, arousal of stereotypes (Nakane 2007), and even stigmatisation (Munro 2003; Piske 2012).

This book aims at shedding light on the acquisition of important prosodic and interactional skills belonging to oral communication which are normally not explicitly addressed in classroom settings. By identifying learners' difficulties in acquiring the native norm, this work lays the groundwork for improving learners' L2 communicative skills and findings are addressed to language pedagogy for future interventions.

In this introduction, I discuss prosodic and interactional competence in relation to the theory and practice of language learning and teaching, providing a framework for their acquisition and the motivation to investigate them in the L2 classroom. Additionally, I outline the book by briefly presenting the three studies conducted and the corpora used throughout.

1.1 Communicative competence in oral performance

Early models of proficiency focussed mainly on grammar, under the influence of the post-Chomskyan interest in formal and abstract aspects of language (Campbell-Larsen 2015; Kramsch 1986). However, it has long been recognised that L2 learners would not be able to take part in spontaneous spoken interactions if they relied solely on their lexicogrammatical competence. Alongside the knowledge of lexis and grammar, so-called *communicative competence* is crucial, described by Hymes (1972: 277) as knowledge about “when to speak, when not, and as to what to talk about and with whom, when, where, in what

manner". In other words, communicative competence is the ability to use language in a socio-culturally appropriate way according to the context and the goal of the communicative exchange. In line with this view, Canale & Swain (1980) argued that socio-cultural and strategic competence are an integral part of communicative competence.

The emphasis on the use of language rather than theoretical thereof knowledge was the origin of Communicative Language Teaching as a methodology for language teaching (CLT, Savignon 1991). Within this framework, the ultimate goal of L2 learning is no longer mastering vocabulary and grammar structures per se, but the ability to apply them appropriately in the communicative context.

This view is embraced by the Common European Framework of Reference for languages (CEFR, Council of Europe 2001), an international standard designed to describe language ability for European languages, which is globally popular especially for its application to English.² The CEFR builds on an action-oriented approach and claims to represent a shift away from the idea of learning as a linear progression through abstract language structures, towards real-life tasks and purposefully selected notions and functions (Piccardo et al. 2018). In other words, the Framework sees language as a means of communication and not as a subject of study. According to these principles, it describes learners as "social agents" (Council of Europe 2001: 9) who, as members of society, constantly need to accomplish communicative tasks in various contexts and under different circumstances. Considering language as a vehicle to socialisation implies attributing high relevance to interaction and putting the co-construction of meaning at the centre of the learning process (Piccardo et al. 2018).

The CEFR provides a schematic description of communicative language competence (Figueras et al. 2009).³ There, communicative competence is depicted as a compound competence formed by several components: linguistic competence, intended as lexical and grammatical knowledge, together with cognitive organisation and accessibility; sociocultural competence, including the knowledge of register appropriateness, degree of formality, rules of politeness and the knowledge of linguistic rituals specific to a community; pragmatic and strategic

²There are various frameworks with the aim of describing language proficiency. Some examples are the American Council on the Teaching of Foreign Languages Proficiency Guidelines (ACTFL), the Canadian Language Benchmarks (CLB) and the Interagency Language Roundtable scale (ILR).

³After a recent revision (consult Piccardo et al. 2018), minor changes were applied to the overall structure. Strategic competence was moved from communicative language competence (left with three main components: linguistic, socio-linguistic and pragmatic competence) and re-categorised as a component of overall proficiency, so that communicative strategies are now specified according to the four different communicative activities.

competence, referring to the functional skills necessary to arrange the message and manage the conversation according to interactional schemata. These components are differently involved in the four language activities learners can be faced with: production, reception, mediation and interaction. Among these language activities, spoken interaction is the most direct and common form of communication in our daily life – nowadays probably comparable only to online written communication in its daily frequency. Therefore, a central role should be ensured for all the skills involved in oral interactive communication in language teaching and testing. Within this framework, the skills investigated in this book – prosodic marking of information status, floor management via turn-taking and backchannel use – are considered necessary for L2 learners to effectively take part in an oral exchange (as demonstrated by the fact that they are listed under “spoken interaction” in the above-mentioned CEFR scheme). In the CEFR, prosodic marking of information status is an ability belonging to “phonological control”, floor management can be synonymously used with “turn taking” and backchannels or feedback expressions are means of “cooperating”, but also contribute to “spoken fluency”.

1.1.1 The theory relating to prosodic competence

The correct pronunciation of a foreign language implies being able to produce both L2 segments, i.e. consonants and vowels, and prosody, i.e. intonation and rhythmic features at a suprasegmental level. Prosody is especially crucial in oral communication, as it plays a fundamental role in conveying and interpreting not only linguistic information, such as discourse chunking, information status (e.g. whether a referent is new, given, or contrastive), or the disambiguation of syntactically ambiguous sentences, but also paralinguistic information indicating a speaker’s identity and attitude.

Starting with the CLT approach, L2 prosody has gradually received more attention in SLA research (Busà 2008; Lengeris 2012) and evidence has been collected regarding its crucial role in the production and perception of meaning. Studies have demonstrated that prosodic features deviating from the native norm can affect listeners’ judgements of accentedness, comprehensibility and intelligibility (Hahn 2004; Jilka 2000; Kang 2010; Kang et al. 2010; Munro & Derwing 2001; Trofimovich & Baker 2006) more than deviations in the production of segments (Derwing et al. 1998; Munro 1995; Munro & Derwing 1999; Gordon & Darcy 2022). A non-target-like use of intonation can even lead to misperception and negative evaluation by native speakers (Munro 2003; Munro & Derwing 2020; LeVelle & Levis 2014).

1.1 Communicative competence in oral performance

In the past, the objective of pronunciation teaching was to eliminate any foreign accent, setting unsustainable learning goals. Nowadays, teaching L2 pronunciation aims at improving learners' communicative effectiveness and success. Exactly for this reason, the description of phonological competence in the early version of the CEFR (Council of Europe 2001) attracted a lot of criticisms for its implicit negative connotation of foreign accents and unrealistic learning goals, as well as for referring to the concepts of stress, intonation, pronunciation, accent and intelligibility in an unclear way (Piccardo & North 2017). Following a recent revision, intelligibility has been adopted as a criterion for defining the progression through proficiency levels, and prosody has been awarded the attention it deserves. The description of prosodic ability in the current version mentions, in particular, the importance of learning how to prosodically highlight newsworthy information:

The focus is on the ability to effectively use prosodic features to convey meaning in an increasingly precise manner. Key concepts operationalised in the scale include the following: control of stress, intonation and/or rhythm; ability to exploit and/or vary stress and intonation to highlight his/her particular message. (Piccardo et al. 2018: 135)

Therefore, the use of prosody to mark information status is officially recognised as an essential skill to be acquired by L2 learners in order to communicate effectively by expressing even the more subtle shades of meaning.

1.1.2 The theory relating to interactional competence

Interactional competence is a concept that was developed to stand in contrast to communicative competence. As pointed out by Young (2014), the position of Hymes (1972) and Canale & Swain (1980) still considers communicative competence as a testable and quantifiable individual characteristic, reflecting the assumption that competence resides in the individual. Even though the notion of communicative competence was useful for moving language teaching away from its narrow focus on lexicogrammar competence, it does not take into account the interactive factor of communication. Therefore, the expression *interactional competence* was coined to refer to what a speaker does when interacting with other individuals, differently from the concept of communicative competence, focussing on what a speaker knows, i.e. the knowledge of an individual (Young 2011). Research in conversation analysis has contributed to this conceptualisation with descriptions of interaction as a joint creation of discourse between interlocutors (Jacoby & Ochs 1995). Moreover, interactional competence has been claimed

1 Introduction

to be context-specific to the extent that it emerges in varied interactive practices to which participants contribute with the appropriate linguistic and pragmatic resources (Hall 1993; He & Young 1998; Hall 1995; Young 2011; as backed up by later research, i.e. Tavakoli 2016; Witton-Davies 2014).

In the CEFR, interaction is described as a language activity in which:

at least two individuals participate in an oral and/or written exchange in which production and reception alternate and may, in fact, overlap in oral communication. Not only may two interlocutors be speaking and yet listening to each other simultaneously. Even where turn-taking is strictly respected, the listener is generally already forecasting the remainder of the speaker's message and preparing a response. Learning to interact thus involves more than learning to receive and to produce utterances (Council of Europe 2001: 14).

In its face-to-face form, interaction requires several skills simultaneously. Learners' productive and receptive skills come into play, as well as additional abilities which allow speakers to monitor the development of the conversation and constantly adjust to it in real-time. To manage the floor, speakers use turn-taking strategies and during the exchange, they co-operate and build the conversation together. Speakers also need to be able to provide feedback, cope with unexpected misunderstandings, ask for clarification, or repair communication breakdowns. Thus, interlocutors constantly evaluate the on-going process to be able to react appropriately. The convergence of all these variables in real time and face-to-face results in a high degree of complexity, to which a second language learner might need to get accustomed before these processes become as automatic as in their first language (L1 – for the concept of automatization in L2 see Kormos 2006). This definition focuses especially on the temporal or rhythmical aspect of face-to-face interaction. Therefore, I will refer to this specific aspect of general interactional competence as *interactional fluency* (which pertains to the fluency of the co-constructed conversational rhythm by the interlocutors, see Peltonen 2024, 2020).

The specific abilities related to turn-taking and feedback signals, which essentially contribute to interactional fluency, are described in the CEFR as follows:

Taking the floor (Turntaking) is concerned with the ability to take the discourse initiative. <...> Key concepts operationalised in the scale include the following: initiating, maintaining and ending conversation; intervening in an existing conversation or discussion, often using a prefabricated expression to do so, or to gain time to think (Piccardo et al. 2018: 140)

Cooperating concerns collaborative discourse moves intended to help a discussion develop. Key concepts operationalised in the scale include the following: confirming comprehension (lower levels); ability to give feedback and relate one's own contribution to that of previous speakers (higher levels); <...> (Piccardo et al. 2018: 101)

Research has also provided experimental evidence for the centrality of these abilities. Turn-taking, to which turn-regulating backchannels contribute, is the foundation of the interaction, representing its empty (i.e. content-free) rhythmic structure, and results from a form of cooperation (see Wehrle 2023 for an extensive discussion). Temporally coordinated collaborative activities, i.e. manual labour, dancing or music-making (see e.g. Hawkins et al. 2013), are greatly attested in human beings. Similarly, tightly synchronised and regulated communicative vocal and/or gestural turn-taking is attested across different species (Pika et al. 2018; Ravignani et al. 2019; Takahashi et al. 2016). However, human turn-taking appears particularly remarkable for several reasons: 1. it is executed with extreme precision and flexibility, 2. it involves the simultaneous prediction, planning and production of utterances, and 3. it is the means through which human language, and to a certain extent human culture, are learned and transmitted (Schegloff 2006). For these reasons, turn-taking and turn-regulating backchannels are two skills to optimally start with when investigating interactional phenomena. This applies even more to SLA research, where very few studies have been carried out on turn-taking and backchannels despite them being primary and perceptually salient abilities in interactions. Indeed, speakers across different native languages have been found to be highly sensitive to turn-timing and tend to avoid very long gaps and overlaps in favour of smooth alternations of turns (Levinson & Torreira 2015; Stivers et al. 2009) because less smooth transitions can cause misunderstandings. For example, an excessively long gap may be interpreted by the recipient as some sort of problem from the interlocutor's side, such as a difficulty in answering affirmatively (Roberts et al. 2011), or uncertainty in reacting to the turn (Levinson 1983). Feedback expressions, or backchannels, can be used to signal comprehension and invite the interlocutor to continue speaking, or start their own turn. Furthermore, interactive practices are often bound to culture, which might represent an additional obstacle for learners. Therefore, mastering turn-taking and backchanneling conventions is necessary for L2 learners to smoothly manage the floor by correctly interpreting and expressing any intention of taking or relinquishing a turn.

1.1.3 The gap between the theory and practice of language teaching and learning

Despite the centrality of spoken interaction being recognised by the CEFR and the CLT approach prevailing in language teaching, prosodic and interactional skills are still neglected in L2 classrooms. Moreover, SLA research extensively reports that learners transfer prosodic (e.g., Rasier & Hiligsmann 2007; Goad & White 2019; Nava & Zubizarreta 2008; Austin et al. 2022; Zhang et al. 2023) and interactional features (e.g., Reinhardt 2022; Sbranna et al. 2020; Handley 2024; Clingwall et al. 2024) from their L1 to their L2, showing that there is a discrepancy between the theory and practice of language teaching and learning, which has an impact on the successful acquisition of the relevant abilities by L2 learners.

One reason for this is the lack of a shared domain between L2 research and pedagogy, which impedes the exchange of knowledge about the results of empirical studies in laboratory settings (prevalent as compared to those in naturalistic settings, see Loewen & Sato 2018) and the concrete application of teaching methods in classroom settings. This gap represents the greatest limitation to the application of research findings to teaching practice. As a result, teachers are left to rely on their own native speaker intuition when it comes to teaching both prosody (Derwing & Munro 2015) and interactional schemata (Campbell-Larsen 2022; Tavakoli & Hunter 2018). This is problematic because teachers generally have only an unconscious competence of these aspects of their native language and culture, so that their intuitions about the pragmatic use of their L1 may not always be accurate (Cohen 2005). Support for this claim comes from teachers reporting that they do not teach pronunciation due to a lack of confidence, skills and knowledge (Macdonald 2002) and from the lack of a common understanding of what fluency is or how it can be fostered in class (Morrison 2018; Tavakoli 2023). Very importantly, many L2 teachers are not native speakers and cannot rely on L1 intuition, in which case a conscious and systematic knowledge of prosody and interactional behaviour is indispensable. For these reasons, it is necessary to provide a solid foundation for applied teaching methods by explicitly addressing research findings on language- and culture-specific aspects of language. Having an empirically-based understanding of these aspects of a language would allow native and non-native teachers to integrate them into their L2 teaching practice.

The aim of this book is to contribute to the empirically-based knowledge of SLA research. Even if these studies are based on a specific population of learners, their findings have broader implications for language research and pedagogy, encouraging replication with other language pairs and the application to teaching practices.

1.2 Outline of the book

The book includes three studies, the topics of which deal with melodic (i.e. prosody) and temporal aspects of L2 interactions (i.e. turn-taking and backchannels which contribute to turn alternation).

Chapter 2 is dedicated to prosodic competence and contains an extensive study on the prosodic marking of information status, which is mentioned in the CEFR as a key ability to master in order to transmit and interpret the partitioning of the message into new or important vs. given or less important information. Previous research comparing West-Germanic languages and Italian has shown that the prosodic marking of information status within noun phrases (NPs) involves a different distribution of pitch accents, i.e. modulations of the fundamental frequency (F0) that are realised on the accented syllable of a word and make it sound prominent (Avesani et al. 2015; Swerts et al. 2002). L1 speakers of West-Germanic languages deaccent post-focal given information, whereas L1 Italians seem to always accent the second word of the noun phrase, regardless of information status. These studies report a discrete classification of accentuation or deaccentuation and a categorical description of the pitch accent type. However, such a categorical description alone can entail missing important information, since speakers may modulate continuous parameters to prosodically mark information structure. I explore these modulations within noun phrases produced by Italian learners of L2 German in the absence of explicit prosodic training using an innovative method based on periodic energy, which reflects pitch perception. Results contrast with previous findings on Italian and show that Italian speakers do differentiate information status within NPs using an information-status-specific timing of F0 movements on the first word, both in production and perception. Comparing the strategies developed in the L2 to learners' native and target languages shows that learners transfer their L1 strategy to the L2, irrespective of their proficiency level, by differentiating information status through a modulation of F0 alignment. However, contrary to productions in their native language and akin to the post-focal given German condition, learners prosodically attenuate the second word of the NPs across the board, probably perceiving the native German deaccentuation as a salient feature of the language. I further show that the method used provides similar or, in some cases, clearer results than well-established measures for acoustic analysis.

Chapter 3 and Chapter 4 are devoted to skills belonging to interactional competence and comprise two studies on turn-taking and backchannels, respectively. These are both underexplored areas of L2 research but core skills that complement each other in spoken interaction.

1 Introduction

Chapter 3 contains the study on turn-taking. This study is inspired by the discrepancy between the theory of the CEFR, rooted in the social paradigm, and the reality of many L2 assessment practices that are still based on grammar and the lexicon. The ability to converse is often only impressionistically evaluated. Thus, I propose a method for the quantification and visualisation of L2 interaction management across different levels of L2 proficiency, suggesting it as a possible starting point for the assessment of L2 interactional competence. The analysis based on the quantification of speech time, silence, overlap, backchannels and dialogue duration shows that the smoothness of the turn-taking system is affected by L2 proficiency, with more overall silence and less speech time for beginners. No effect of lexical competence was found on learners' conversational patterns, which indicates the independence of the two types of competence. With the same method, I also propose a preliminary cross-linguistic comparison between native Italian and German interactions (learners' native and target language respectively), pointing out that Germans might have a more careful approach to task completion, with longer speech time and dialogue duration. Thus, the method proposed appears to be useful to capture differences in turn-taking practices across L2 proficiency levels, as well as languages.

Chapter 4 is dedicated to backchannels. This study is complementary to the one on turn-taking, as I take into account backchannels with turn-regulating functions, the production and interpretation of which contributes to a smooth turn alternation. Previous research reports that backchannels have positive social implications, signalling engagement in a conversation. Nevertheless, their language-specificity represents an obstacle for L2 learners, who might face miscommunications and misperceptions in case of an inappropriate use of backchannels in intercultural conversations. This study aims at investigating whether learners acquire target-like backchannelling behaviour even in the absence of explicit instructions, since this aspect of interactional competence is generally not thematized in L2 classroom settings. I present an in-depth study across native Italian and German and in L2 German, accounting for several backchannel features. In both L1s, findings show a complex, non-arbitrary mapping between lexical type, function and intonation. Backchannel frequency was found to be similar across languages, while backchannel duration is longer in German. The learners' proficiency seems to only play a role in the lexical choice of backchannels, whilst dyad-specific patterns appear to account for the frequency and duration of backchannels in the L2 better than proficiency.

Chapter 5 concludes the book. It begins with a summary of the three studies, followed by a discussion of the findings in the context of second language

research and teaching. Finally, the limitations of the present investigations are examined and concordant directions for future research are suggested.

1.3 Corpora

Since participants and the data collection method are shared between the three studies, all information relative to these aspects will be given in this introductory chapter and referred to throughout the book.

1.3.1 Participants

40 Italian native speakers learning German, 14 Italian monolingual⁴ native speakers and 18 German native speakers were recorded while performing dyadic task-based conversations.

All Italian speakers had grown up in the dialectal area of Naples with parents of the same origin. Thus, variation resulting from their native linguistic substratum can be ruled out. Learners were either students at the Goethe Institute in Naples (aged between 23 and 65, mean = 33; median = 30; SD = 12.29; 6 females, 4 males), or at the Department of Literary, Linguistic and Comparative Studies (It.: Dipartimento di Studi Letterari, Linguistici e Comparati) at the L'Orientale University of Naples (aged between 19 and 25, mean = 21; median = 20; SD = 1.2; 27 females, 3 males), with German as a foreign language as one of their main subjects.⁵ Italian monolingual speakers were students of subjects other than foreign languages at different universities in Naples (aged between 19 and 24, mean = 22; median = 22; SD = 1.80; 10 females, 4 males). They reported to have been exposed to English, or in some cases French and Spanish, in school. However, at the moment of the recording they only had a passive reminiscence

⁴By “monolingual” I mean that, despite having been exposed to some foreign languages in school, at the moment of the recording they were neither proficient in nor active speakers of any foreign language, due to the lack of exposure and exercise, as opposed to the learners of German.

⁵Twenty-four learners had spent time in a German-speaking country for varying durations (ranging from one to ten months), either for a short language course or an exchange program at a partner university. However, the impact of time spent abroad is neither easily quantifiable nor uniform across individuals and depends on factors such as the amount of exposure to and use of the foreign language. For instance, some exchange students may struggle to establish regular contact with locals or choose not to enrol in a German language course. That said, because immersion in a foreign language generally contributes to overall proficiency, I did not treat this variable separately.

1 Introduction

of these languages, thus at proficiency levels which should not be assumed to affect their native language.

Native German participants came from different dialect areas,⁶ but had been living in Cologne for at least three years at the moment of the recording, and were students of subjects other than languages at the University of Cologne (aged between 22 and 27, mean = 24; median = 24.5; SD = 2.49; 11 females, 7 males).

No subject reported to have ever received specific phonetic and/or interactional training, nor to suffer from any speech or hearing problem. As colours were involved in one elicitation game, I also made sure that no participant was colour-blind.

Learners' proficiency levels were established on the basis of the language courses they were attending at the time of the recordings and ranged from A2 to C1 CEFR levels, in which the notations "A", "B", "C" for proficiency levels correspond to beginner, intermediate and advanced levels of competence in German. The proficiency groups resulting from the data collection were unbalanced in number, with only six A- and four C-level learners. Thus, for the sake of a more reliable statistical analysis, learners were recategorised into two proficiency groups, each with a similar number of participants. Nevertheless, speaker- or dyad-specific variability is discussed in the three studies whenever it was found to be relevant for the interpretation of the results on the group level. I defined learners with A2 and B1 levels as beginners and learners with B2 and C1 levels as advanced. This division is not only based on the midpoint of the CEFR proficiency scale, but also on the structure of the reference levels themselves. Indeed, the gap between the abilities required by the B1 (also called "Threshold") and B2 (also called "Vantage") levels is greater than the one between C1 and B2, which makes it a suitable demarcation line for recategorising proficiency levels into two groups only. All dyads of learners, alongside their corresponding CEFR and recategorised levels of L2 German, are listed in Table 1.1.⁷

⁶Due to the limitations imposed by the pandemic it was not possible to strictly select German native speakers from the same dialect area. In particular, they were born in places above the Benrath line (Wenker 1877) and between the latter and the Speyer line (Paul 2013). Thirteen speakers were from North Rhine-Westphalia, two from Lower Saxony and three from Hesse. However, nobody reported a mastery of the dialectal variety of their place of origin.

⁷Due to participant availability, dyads 11 to 14 are mixed, i.e. they are composed of one learner at B1 level and one learner at B2 level. Mixed dyads were recategorised based on the proficiency level of the instruction giver (a role in the Map Task, which generally leads the conversation), except for ME. I did so as the giver of this dyad was following a B1.2 course, whereby the notation ".2" is used to describe CEFR proficiency levels with a greater degree of detail and stands for an advanced mastering of the level it accompanies. Moreover, this participant had recently spent 9 months in Germany. As a result, his performance, i.e. fluency, vocabulary, grammar structures, was comparable to those of B2 learners.

Table 1.1: Proficiency level of learner dyads.

Dyad ID	CEFR level	Recategorised level
1 IF	A1	Beginner
2 CV	A2	Beginner
3 AR	A2	Beginner
4 RM	B1	Beginner
5 CC	B1	Beginner
6 AA	B1	Beginner
7 AC	B1	Beginner
8 AN	B1	Beginner
9 GS	B1	Beginner
10 GA	B1	Beginner
11 CA	B1-B2	Advanced
12 RS	B1-B2	Advanced
13 CR	B1-B2	Advanced
14 ME	B1-B2	Advanced
15 AB	B2	Advanced
16 CE	B2	Advanced
17 MA	B2	Advanced
18 RC	B2	Advanced
19 FF	C1	Advanced
20 BS	C1	Advanced

1.3.2 Data collection

Mono recordings of uncompressed WAV files at 44.1 kHz sample-rate and 16-bit depth were collected using headset microphones (AKG C 544 L) connected through an audio interface (Alesis iO2 Express). Each recording session included two conversational games: a semi-spontaneous interactive board game to elicit the prosodic marking of information status, and, subsequently, a conversational game, the Map Task (Anderson et al. 1991). The two conversational games are explained in the respective studies.

Three groups of participants took part in the experiment and were recorded in pairs: Italian learners of German, who were recorded in their native Italian and L2 German, Italian monolingual native speakers, and German monolingual native speakers. Italian learners of German could self-select their partner for the recordings with the only requirement that they had the same or a similar proficiency

1 Introduction

level of L2 German, i.e. they were classmates at the university or the Goethe Institute. L1 Italian and German monolingual speakers were also preferably matched following the criterion of self-selection and, in a minority of cases, based on the participants' schedule. The Italian paired learners were mostly classmates and already knew each other before the recording session; Italian monolingual pairs were mostly matched according to their availability and the majority of them did not know each other before the recording, while native German pairs were all self-selected and speakers already knew each other.

In each recording session, two participants sat at two opposite sides of a table. Eye-contact and signal interference between the two microphones were prevented using an acoustic insulator dividing panel which was opaque, making it impossible for participants to see each other and the other person's materials in order to maximise oral communication. Controlling this parameter was particularly crucial for the studies focussing on verbal interactional skills of learners, since backchannels and turn-taking can be conveyed via non-verbal cues like eye gaze, head movements, gestures and facial expressions (which are visual communicative channels in oral face-to-face interaction, see Kim et al. 2024; McDonough et al. 2024, but not the focus of this research). All pairs first played the board game and then performed the Map Task. Both tasks were introduced by written instruction and participants were given the chance to ask clarification questions before the beginning of the tasks. In the case of learners, the tasks were first completed in Italian and then repeated in German. This fixed order was chosen to prevent potential issues arising from misunderstandings of instructions in a foreign language and to avoid introducing additional variability due to task order, which might have obscured proficiency effects. Before carrying out the same tasks in their second language, learners watched video instructions explained by a German native speaker to help them get into the language mode and reduce L1 bias. Finally, all speakers were provided with a sociolinguistic questionnaire. Learners also completed the German version of LexTALE (Lemhöfer & Broersma 2012), an online test for L2 lexical competence.

Italian participants were recorded at the Goethe Institute in Naples. The recording session lasted approximately 90 minutes for learners since they performed the tasks in both their L1 and L2, whereas Italian monolinguals took 45 minutes. German native speakers were recorded at the University of Cologne and recording sessions lasted 45 minutes.

Finally, some ethical aspects are in order. The study was conducted in compliance with ethical standards for research involving human participants. All participants were provided with and signed an informed consent prior to their

participation in the study, acknowledging their voluntary involvement and understanding of the research purpose. The consent process ensured participants were aware of their rights to withdraw at any time without repercussions. Participants received financial compensation for their time. Data processing adhered to the principles of GDPR compliance, ensuring the protection and confidentiality of personal data. Any identifying information was anonymised during data analysis and storage.

1.3.3 Data use across the studies

For each of the two games used for data elicitation, the resulting corpus consists of twenty dialogues in Italian and twenty dialogues in German as an L2 by Italian learners, seven dialogues in Italian L1 by monolinguals and nine dialogues by German native speakers.

Considering that during classes, exposure to the target prosody is proportionally greater than exposure to culture-specific interaction mechanisms, prosodic features might be implicitly acquired by repeatedly listening to the native teacher and recorded listening exercises, but no spontaneous dyadic interaction between two natives can be observed in a classroom setting (apart from exceptional cases of guests and exchange students) for implicit learning to take place. Therefore, in Chapter 2 dedicated to prosodic competence, the Italian monolingual group will be used as a control group for learners' Italian prosodic realisations since the learners' native prosody might have been influenced by the German-language dominant setting of the recordings, i.e. the Goethe Institute.

In Chapter 3 and Chapter 4, dealing with interactional competence, I will not take into account the Italian monolingual control group. It is reasonable to suppose that learners' L2 style of interaction will not be different from their native one, since spoken interaction in L2 classrooms is mostly exercised among students themselves, and there is not enough foreign exposure to favour the acquisition of a target-like conversational style.

The test for L2 lexical competence is discussed in Chapter 3, in which I compare lexical competence to overall communicative competence.

2 Prosodic marking of information status

This chapter contains an extensive experimental study dedicated to a specific aspect of L2 prosodic competence, the prosodic marking of information status.¹ This ability is the core of prosodic competence (a view also shared by the CEFR, Council of Europe 2020), as it allows learners to highlight the relevant parts of their message, guiding the listener in the correct interpretation of their speech.

The objective of this study is to enrich the existing body of knowledge on L2 prosodic marking of information status by providing new evidence based on a continuous approach and unveiling phenomena which had not been observed yet. Previous research has largely relied on categorical approaches, which may not fully capture the dynamic nature of evolving systems such as L2s. Therefore, in this study I make use of an innovative method for phonetic analysis. To motivate and validate this new method, I also provide an alternative analysis using established measures to compare with and propose a possible categorical interpretation of the results, showing that methodological choices can influence our conclusions about SLA. In light of these reflections, I finally provide suggestions for future prosodic analyses and pedagogical applications in the field of SLA research.

Given the unavoidable technicalities present in this chapter, each section contains one or more summaries reporting the main findings in simple and concise form. These summaries can be read independently and sequentially to get a quick overview of the results.

¹The analysis of L1 Italian speech was previously published in Sbranna et al. (2023). The analysis of L2 German speech, along with its comparison to learners' native and target languages, was previously published in Sbranna, Albert, et al. (2025). This chapter is enriched with additional analyses and connections between individual sections, providing a comprehensive contribution to methodological approaches in phonetic SLA research, as well as insights into the pedagogical implications of the specific results.

2.1 Background

2.1.1 Information structure

To ensure the correct interpretation of their communicative intentions, speakers distribute information throughout discourse. The term “information structure” (IS) was introduced to describe this phenomenon by Halliday (1967). Halliday defines IS as the partitioning of a discourse into “information units”, which are distinct from syntactic constituents, but correlated with “tone groups”, i.e. with intonational phrasing. Another definition was offered by Chafe (1976), who described IS as information packaging, responding to the immediate and temporary communicative needs of speakers. Information packaging operates on the basis of interlocutors’ “common ground” (Chafe 1976; Karttunen 1974; Krifka 2008; Stalnaker 1974), which has been defined as “shared” (Clark & Haviland 1977), “common” (Lewis & Stechow 1979) or “mutual” (Schiffer 1972) knowledge that the parties to a conversation recognise as shared (Stalnaker 2002). In other words, speakers organise their discourse based on the knowledge they assume their interlocutor is familiar with (Prince 1981).

In Halliday’s theory, the IS of an utterance is based on two components, one more informative and one less informative, which are marked by linguistic means. Many terms have been used to describe this informational contrast, and there is no consensus on the different categories of information structure (Büring 2007; Von Heusinger 1999). In the present work, two terminological oppositions are relevant and will be described in detail, *new* vs. *given* and *focus* vs. *background*.

The terms “new” and “given” refer to the mental representations of discourse referents at two different levels (Chafe 1994; Lambrecht 1994): identifiability, i.e. the assumed listener’s knowledge of that referent (Prince 1981); and degree of activation, i.e. the assumed listener’s consciousness of that referent at a certain moment in time during the discourse (Chafe 1994). A constituent is given if it is present in the immediate common ground (Krifka 2008), whereas in the converse case it is new. While the distinction between new and given was regarded as a dichotomy in earlier studies (Brazil et al. 1980; Halliday 1967), it has since been constructed as a continuum along which different degrees of givenness reflect the degree of cognitive accessibility of the information at a certain moment in time (Baumann 2005; Baumann & Grice 2006; Baumann & Riester 2012; Chafe 1994; Prince 1981).

Focus-background structure refers to the pragmatic partitioning of an utterance into informative and less informative parts based on speaker’s *intentions*.

In particular, the term “focus” refers to the element of the sentence that conveys newsworthy information, while “background” refers to less newsworthy elements (Halliday 1967; Kuno 1978). The domain of focus is denoted as *broad* if it extends to a whole constituent or sentence and does not pragmatically single out a specific element (Gussenhoven 1983; Ladd 2008). In contrast, if the domain of focus corresponds only to a selected portion of the utterance, the type of focus is *narrow* (Ladd 1980). Narrow focus can also be contrastive or corrective, where the relative referent either contrasts with a referent in the previous context or serves as a correction (Gussenhoven 2008; Klassen et al. 2016; Krifka 2008; Vander Klok et al. 2018).

These two dimensions of IS, i.e. new-given and focus-background, are separate dimensions, but can intersect. Whilst referential givenness is orthogonal to focus, in that both given and new elements (representing the two poles on the givenness scale) can be focused, focus and newness are correlated (Kügler & Calhoun 2020). The present study is concerned with only two degrees of givenness: a) referentially and lexically given (with the referent being part of the immediately preceding context); b) referentially and lexically new (with the referent not being present in the immediately preceding context). Moreover, in this study newness correlates with contrastive focus (the new element is a contextually identifiable alternative of an element of the same category in the immediately preceding context). With this in mind, in the next section I will review previous research investigating prosodic encoding of both givenness and focus, both of which are relevant for the current study.

2.1.2 Prosodic encoding of information structure

Languages differ in the linguistic means used to mark information structure – e.g. prosody, syntax, and lexical markers. In languages that make use of prosodic means to mark prominence, speakers prosodically attenuate shared or unimportant information and highlight new or important information. The highlighting function of prosody consists in making one element prominent compared to its neighbours, namely to make it “stand out” from its context by virtue of its acoustic characteristics (Cangemi & Baumann 2020; Terken 1991). Different phonological and phonetic properties can contribute to prosodic prominence, e.g. accentuation, pitch accent type, phrase boundaries, or pitch register (for a review, see Kügler & Calhoun 2020).

In many stress-based systems, accentuation is used to prosodically realise prominence, reflected in a bundle of phonetic cues corresponding to change in F0, local increase in duration, enhanced overall energy and spectral properties

on and around the stressed syllable (Baumann & Winter 2018; Campbell 1995; D’Imperio 2000; Heldner 2003; Hermes & Rump 1994; Kochanski et al. 2005; Kügler & Calhoun 2020; Sluijter & Van Heuven 1996; Turk 2012). By contrast, deaccentuation is defined as the absence of accentuation where an accent would have been expected in a comparable all-new utterance, usually in final position (Cruttenden 1997; Ladd 1980), reflected in prosodic attenuation.

The majority of the studies on stress-based prosodic encoding of IS have been carried out within the Autosegmental-Metrical (AM) phonology framework (Beckman et al. 2005; Ladd 2008; Pierrehumbert 1980) and predominantly on Germanic and Romance languages. According to AM, the main stressed syllable of the head of a prosodic unit is the most prominent one, to which the nuclear pitch accent is assigned – by default the final fully-fledged pitch accent of the prosodic unit. Therefore, focus is considered to be only indirectly marked by phonetic cues: these cues would, in fact, mark nuclear pitch accents, which, in turn, mark focus (Büring 2016; Calhoun 2010; Ladd 2008; Selkirk 1995). As a consequence, post-nuclear phonological heads are either deaccented, i.e. not associated with a pitch accent, or they present an accent with a compressed pitch register (Kügler & Féry 2017), with exceptions occurring only in specific discourse contexts.

However, even within stress-based systems, languages differ in the extent to which this strategy is applied and aligned with syntactical and lexical means. For example, some Romance languages (also stress-based systems) have been claimed to follow the default in situ stress-based pattern (i.e. the nuclear accent can be in any position in the utterance), while others appear to use in situ stress-based marking only in specific discourse contexts, such as contrastive focus, and no post-nuclear deaccentuation (e.g. Italian, Catalan and Madrid Spanish: Frota & Prieto 2015; Vanrell & Fernández-Soriano 2018).

In the following section, I will review the different methodological approaches used to investigate the phonetics and phonology of prosodic marking of information status.

2.1.3 Categorical and continuous approaches

The long tradition of studies on West Germanic languages has yielded evidence for a close relationship between prosodic accentuation and information structure. In particular, focus and newness are considered to be prosodically indicated by the presence of a pitch accent, usually a nuclear one, with background and given elements usually deaccented (Cruttenden 2006; Féry & Samek-Lodovici 2006; Halliday 1967; Ladd 1996; Terken 1984). The extensive evidence collected in

favour of consistent associations between certain intonation patterns and their functions has often led to the simplistic assumption of a binary distinction between new and given information, and of a one-to-one relationship between accentuation and information status (Halliday 1967).

The earlier studies on the intonation of West-Germanic languages influenced successive research questions on Romance languages. A clear example is the use of the term “re-accentuation” of given elements in early research on Romance languages, to contrast with the “deaccentuation” in West Germanic languages (Cruttenden 1993; Swerts et al. 1999), implying that no accent is in some sense the subtraction of an accent that is assigned by default. However, Cruttenden (1993, 1997) himself observed that these consistent associations are, in fact, preferred patterns and that languages generally allow for alternative ways of expressing the same function. Nonetheless, the West-Germanic pattern is seen as the default.

Successive experimental evidence has shown that not only new but also given information can be accented (Bard & Aylett 1999; Riester & Piontek 2015; Schweitzer et al. 2009; Terken & Hirschberg 1994) and that the relation between pragmatic functions and accentuation – and different pitch accent categories – is rather more complex (Grice et al. 2017; Krahmer & Swerts 2001; Mücke & Grice 2014). For example, Chodroff & Cole (2019) found in a study on American English that given information was mostly unaccented or conveyed through low pitch accents (L*), while contrastive information was mostly marked by high (H*) and rising accents (L+H*). However, the opposite relation was found as well: given items realised with high or rising accents and new and contrastive items realised with a low accent or no accent. Similar findings across languages led to the analysis of a probabilistic relation between form and function (for English: Calhoun 2010; Chodroff et al. 2019; Ito et al. 2004; Yoon 2010; for German: Baumann & Riester 2013; De Ruiter 2015; Kurumada & Roettger 2022; Röhr & Baumann 2010).

The same research development from a rigid categorisation to a probabilistic mapping took place for Romance languages. Traditional focal typology used to categorise Romance languages as “non-plastic” languages, with rigid prominence patterns and obligatory word order modifications for expressing focus and information status, in contrast to “plastic” languages, such as West Germanic languages, which can flexibly modify prominence patterns in a sentence to accent focused or new information (Vallduví 1991). Later research attested the use of both strategies in some Romance languages (e.g., Italian) and described it in terms of “preference”, even if these studies still employ a binary distinction between languages with deaccentuation for marking given elements, on the one

hand, and languages with marked syntactical structures but limited deaccentuation on the other (Cruttenden 1993; Ladd 1996, 2008). A different proposal is based on evidence from Spanish, Italian and English: Face & D’Imperio (2005) show that differences across West-Germanic and Romance languages relate to their distribution, with English only rarely using word order, as compared to Spanish and Italian. Further, there is an interaction between the two strategies, with Spanish using either word order or intonation; and Italian requiring both, i.e. word order and intonation in utterance-final position but intonation only in utterance-medial position. The resulting complex picture cannot be accounted for with a simplistic binary distinction. Instead, a typological continuum with syntax (referred to as word order) and intonation at two opposite poles is proposed.

Recent research has raised awareness of the limitations of a purely categorical approach and the need to account for within- and cross-language variability. Cangemi & Grice (2016) point out that a distributional approach to phonological categories can better account for the phonetic variability found in encoding. They compare categories to clusters in a multidimensional space which include substance – in Saussurian terms (De Saussure 1916) – along with form and function. Clusters can vary in their compactness and internal structure, possibly presenting sub-clusters (different variants to express the same function). Such an approach considers variability as an integral part of the categories and, consequently, as a source of information about their internal structure. As a result, related studies in intonational phonology have used approaches that explore the continuous dimensions of intonational categories (Cangemi & Grice 2016; Grice et al. 2017; Roessig et al. 2019).

Findings on gradient and probabilistic mapping of form to function have generated a growing interest in the description of phonological pitch accent categories in terms of the acoustic correlates that are used to encode them in both production and perception.

Many studies have focussed on the relevant continuous modulations of the magnitude and the timing of events in the fundamental frequency contour with respect to landmarks in the segmental string (see Section 2.8.1 for further details on the segmental anchoring hypothesis and autosegmental-metrical model) and, interestingly, found that the modulation of these phonetic parameters accounts for both variability in the mapping between form and function and inter- and intra-individual variability (Cangemi et al. 2016; Cangemi & Grice 2016; Cangemi et al. 2015; Grice et al. 2017; Mücke & Grice 2014). For example, Grice et al. (2017) investigated the mapping of different focal structures onto their prosodic realisation both in terms of pitch accent types and of continuous phonetic parameters

contributing to pitch accent categorisation, i.e. F0 peak alignment, target height and tonal onglide. Results of the GToBI (German Tones and Break Indices: Grice et al. 2005) analysis showed that not all speakers use the same accent type to express the same function, but the distribution of continuous parameters revealed systematicity in the patterns for all speakers. For some speakers, the difference in modulation of these continuous parameters was clear enough to cause a shift in the transcription of accent type, while for other speakers the modulation was more subtle and did not lead to the transcription of a different accent category.

Perception studies further validate these results, as listeners accurately perceive the modulation of these parameters in marking a specific function, even when different speakers use different strategies (Cangemi et al. 2015; Grice et al. 2017). Further evidence suggests that speakers and listeners differently weigh multiple dimensions to account for the same category (Cangemi et al. 2015; Grice et al. 2017; Niebuhr et al. 2011). Thus, the subtle modulation of different phonetic parameters plays a distinctive role in both the expression and identification of functions, contradicting strictly categorical accounts.

In the next sections, I will review the state of the art on the prosodic marking of information status in both Italian and German (the native and the target language of the learners investigated in this study), and finally review the L2 literature, including a pioneering comparative study on Italian and German as L1 and L2.

2.1.4 Prosodic marking of information structure in Italian

Research on the prosodic marking of information status in Italian offers a fragmentary picture due to differences in experimental design, making a comparison of results difficult. The picture is further complicated by the fact that there is no consensus about a standard intonation and rhythm in Italian, since it presents a great deal of variation across dialectological areas (Bertinetto & Loporcaro 2005; Canepari 1980; Giordano 2006; Lepschy & Lepschy 1977; Magno-Caldognetto et al. 1978; Savino 2012; for a comprehensive review see Gili Fivela et al. 2015; Vietti 2019).

Some studies have focused on the prosodic expression of givenness in Italian in light of the phenomenon of deaccentuation that is widely attested in West Germanic languages. These studies are mostly based on the observation that Italian seems to allow deaccentuation of entire syntactic constituents, i.e. full clauses or noun phrases, but blocks deaccentuation within syntactic constituents (Ladd 1996).

Evidence shows that Italian can mark contrastive focus with an accent and then deaccent the following constituent in sentence-length utterances, as in West

Germanic languages. In northern Italian varieties, an investigation of different focus structures in sentences composed of two phrases – a noun and a verb phrase – revealed that post-focal falling contours have reduced F0 range and duration as compared to final falls in a neutral condition (Farnetani & Zmarich 1997). In Tuscan Italian (Avesani et al. 1995; Hirschberg & Avesani 1997) and Neapolitan Italian (D’Imperio & House 1997), a low and flat F0 contour with no evidence of F0 movement on post-focal given elements was also found in both full clauses and simple phrases. Indeed, Avesani et al. (2015) argue that Ladd’s observation should rather be interpreted as an indication that deaccenting of given elements can, in fact, occur in Italian.

However, there are also reports claiming the opposite. In Tuscan Italian broadcast speech, textually given elements were found to be accented with an L* irrespective of their grammatical function and position in the sentence (Avesani 1997). Similarly, in Roman Italian task-oriented dialogues, most coreferential given elements were found to be accented irrespective of their position in the discourse and prosodic unit, and only a few cases of post-focal deaccentuation are reported in sentences with fronted foci (Avesani & Vayra 2005). No relation was found between a specific pitch accent and givenness. Further investigations on Tuscan Italian have found that post-focal given elements occurring in a metrically strong position, i.e. as head of the prosodic domain, show a clear increase in duration, as well as in formant and spectral emphasis, but no F0 movement (Bocci 2013; Bocci & Avesani 2011, Avesani et al. 2015). Thus, it has been suggested that the low and flat F0 contour found on post-focal given elements should be interpreted as an L* pitch accent, rather than as deaccentuation (Bocci 2013).

The presence of an accent in the flat post-focal region is also suggested by another study on the production and perception of focus in short Neapolitan Italian sentences (D’Imperio 2002). The author hypothesises that the flat postfocal region is a compressed, downstepped version of the non-salient H+L* phrase accent, which would also explain why it lacks salience in perception.

These studies, despite providing important experimental evidence for the prosodic expression of givenness in Italian, do not yield a clear picture, owing to differences in their design and materials (e.g., the variety of Italian under investigation, size of the data sample, data collection method, speech style analysed, and analysis).

2.1.5 Prosodic marking of information structure in German

In German, prosody is the main linguistic marker of information structure. The traditional assumption that new elements are accented and given ones deac-

cented (Allerton 1978; Cruttenden 2006) has been relativised by several studies (Baumann 2005; Baumann & Hadelich 2003; Kügler & Féry 2017; Wagner 1999). These investigations still provide evidence that deaccentuation is the most appropriate and common way to mark givenness but show that different options are available. For example, Baumann & Riester (2013) found fewer deaccentuation cases than expected from the literature in a study on both read and spontaneous speech, and claim that there is a differential probability for an item to receive an accent. Therefore, the authors suggest a gradient scale of prosodic prominence, realised through a range of different accent types (including deaccentuation), mapping onto different degrees of activation of a referent (Baumann 2006; Baumann & Riester 2013). In turn, different pitch accent types are realised through distinct modulations of continuous phonetic parameters.

The central phonetic cue used in German to mark information status is pitch excursion. Féry & Kügler (2008) found a correspondence between information status and tonal scaling, with narrow focus raising the high tones of pitch accents and givenness lowering them in prenuclear position and cancelling them out in postnuclear position. Similar results are also reported for analyses of several phonetic cues contributing to prosodic marking of focus (e.g., pitch excursion, peak position, duration and accent type in Baumann et al. 2006; accent type, duration and articulatory gestures in Hermes et al. 2008). Peak position appears to play a role as well. Kohler (1991) found that different accent contours were perceived as corresponding to different meanings, i.e. late peaks (described as $L+H^*/L^*+H$) for emphasis or contrast, medial peaks (H^*) for new information and early peaks ($H+L^*/H+!H^*$) for accessible or given information. In particular, a categorical distinction was only found between early and medial peaks, whereas there was a gradient difference between medial and late peaks.

These studies provide a much clearer picture of prosodic marking of information status and focus in German as compared to the state of the art on Italian, clearly showing that there is a probabilistic and gradient relation between information status and accent type (including the absence of an accent).

2.1.6 Information structure and prosody in L2

Since languages differ in how prosody is used to mark IS, one interesting research question concerns how L2 learners acquire and develop these patterns. To date, there are very few studies on the prosodic encoding of information structure in interlanguages, i.e. a system distinct from both the native and the target languages (Selinker 1972), although it has important consequences for successful communication and the potential for research and educational application is

therefore considerable. The little evidence collected so far is mostly on English as L2 and results are contradictory with regard to the effects of proficiency.

Learners' prosodic encoding of IS has been found to differ from English native speech both in phonetic and phonological terms. From a phonetic point of view, differences in the use of peak alignment, pitch height and pitch movement have been found. For example, a delayed pitch peak on new information was found in Korean learners of English (Trofimovich & Baker 2006), probably due to an influence of L1 pitch alignment patterns. The interlanguage of Mandarin Chinese learners of English, instead, presented less difference in pitch excursion across new and given items compared to L1 English speakers (Wennerstrom 1998). The same result is reported for Spanish learners of English, whose pitch range on focused constituents is narrower, without a clear differentiation from the adjacent syllables. Moreover, these learners were also found to produce a falling contour on both new and given elements, while native speakers differentiated the information status using a fall and a low rise respectively (Verdugo 2003). A study on Malay learners of English showed that the phonetic details of L2 rises resembled those in the speakers' L1 (Gut & Pillai 2014). From a phonological point of view, there seem to be a common tendency for non-native speakers to overaccentuate, regardless of IS (Austrian learners in Grosser 1997; Spanish learners in Verdugo 2003; learners of various L1 backgrounds in Gut 2009; Malay learners in Gut & Pillai 2014). In interpreting this result, it is important to keep in mind that the target language is always English, in which accentuation is used to mark IS, whereas learners' native languages might have a looser relation between accentuation and IS, and might follow other criteria (e.g. phonological ones) to distribute accents in the utterance, leading to the accentuation of given elements in the L2. This is the case for Spanish and Malay learners accentuating given elements in final position, where final position is taken to be the default, or unmarked case, i.e. "the pattern that is chosen when there is no compelling grammatical or contextual reason to choose some other" (Ladd 2008: 223).

In a cross-linguistic study on a language pair not including English, i.e. Dutch and French (Rasier et al. 2010), it was found that not all learners tend to overaccentuate, and that both Dutch learners of French and French learners of Dutch transfer their L1 features in prosodically encoding the IS of noun phrases. Specifically, Dutch learners tended to use the less common French "focus accent" (more similar to their own L1), in which only one element of the noun phrase is accented, and never the more common French "bridge accent", in which both elements of the noun phrase are accented. French learners also applied their own L1 strategy to the L2 in not deaccenting contextually given information. The difference between learner groups is explained by the Markedness Differential Hypothesis

(Eckman 1977) according to which marked structures (such as accentuation according to pragmatic contexts) are more difficult to learn than unmarked ones (such as accentuation in the default condition), so that Dutch learners have an advantage over French learners in this context. Another study conducted on languages other than English reached similar results (German and Dutch learners of Italian in Turco et al. 2015). While in L1 Italian polarity contrasts was marked through a verum-focus accent (i.e., accent placed on the finite verb in verb second position used to emphasise the truth of the sentence as in Höhle 1992) in a minority of cases, German learners of L2 Italian produced more verum-focus accents and transferred their L1 phonetic implementation by often deaccenting post-focal constituents instead of using post-nuclear pitch accents. Transfer was also found for Dutch learners of L2 Italian who preferred lexical markers as in their L1. From these findings based on L2s other than English, it appears that the influence of the L1 may better explain the consistent overaccentuation found in L2 studies on the prosodic encoding of IS (although there are few studies directly comparing learners' L2 to their L1 productions to provide evidence for transfer) compared to the proposal that overaccentuation might be a universal tendency (Gut et al. 2013; Gut & Pillai 2014).

Results relative to the role of proficiency are contradictory. Some studies show that transfer of L1 features tends to reduce as language proficiency increases. As an example, in a study comparing speakers of L1 Zulu with their L2 English, it was found that beginners did not mark information status within noun phrases through accents for contrastive or corrective focus as in their L1, while advanced learners tended to appropriately associate accentuation and information status in contrastive focus (Swerts & Zerbian 2010). Likewise, advanced Japanese learners of English were able to map given information to deaccentuation and contrastive information to an L+H* accent in the same way as native English speakers in rating and production tasks. Differently, less proficient L2 learners associated given information with deaccentuation and contrastive information with L+H* in the rating task only (Takeda 2018). In contrast to these studies, some other studies do not find an improvement in intonational competence corresponding to higher proficiency (Bi 2008; Verdugo 2003), so that the role of proficiency in the learning process remains unclear.

2.1.7 A comparative research programme on Italian and German as L1 and L2

One pioneering research programme has allowed for the comparison of Italian and West-Germanic languages with an experimental design that brought to light

differences ascribable to language structure: Swerts et al. (2002) and Avesani et al. (2013, 2015) investigated the prosodic marking of information status in Tuscan Italian and Dutch, and in Tuscan Italian and German, respectively. Their findings support Ladd's (1996) observation that Italian strongly disfavours deaccentuation within noun phrases as opposed to West-Germanic languages.

Swerts et al. (2002) used a card game to semi-spontaneously elicit noun phrases composed of two words (a noun and an adjective), which could be new, given or contrastive according to the context. For Italian, they report an F0 excursion on both words with a hat pattern stretching over the entire noun phrase regardless of the varying information structures, while in Dutch the F0 excursion correspond to the new element only. It was concluded that Italian fails to deaccent post-focal given elements within noun phrases. A following perception experiment reinforced this finding: Italian listeners could not reliably reconstruct the context in which the noun phrases were produced when listening to them in isolation. This result was replicated in a second perception experiment with the same data (Krahmer & Swerts 2008), which suggests that these utterances lack any other prosodic cues upon which listeners can rely to identify their information structure.

Avesani et al. (2013, 2015) successfully replicated the production experiment by Swerts et al. (2002), reporting that in Italian the second word of the noun phrase is always accented, independently of its pragmatic status, whereas the first word can lack an accent in some cases. Interestingly, a range of pitch accents was found for both the first and second words – H*, H+L* and L+H*, but not including the L* found in previous studies on Tuscan Italian (Bocci 2013). The explanation given is that, in Italian, phonological constraints override the mapping between prosody and pragmatic functions, such as focus or information status. In detail, it is argued that the two words of the NP constitute an intonational phrase, whose metrical head at the rightmost position is the stressed syllable of the second word. The first word, being in pre-nuclear position, can optionally be accented, but does not have to be. In contrast, the metrical head (the second word of the NP) has to bear the nuclear accent and the presence of the nuclear accent in the rightmost strong metrical position cannot be overridden by syntactic or pragmatic requirements (Avesani et al. 2015). In German, according to this view, deaccenting of the strongest metrical position of the intonational phrase is allowed and occurs when required by the information structure (i.e. in the case of given items).

Avesani et al. (2013, 2015) also extended the investigation to L2 German spoken by Italian learners and L2 Italian spoken by German learners, with the same proficiency level in their respective L2s. They found that Germans successfully

reproduced the Italian accentuation pattern, but Italians did not, as only a trace of deaccentuation of given post-focal elements was found in their L2 German (17% of the cases; statistically significant). The authors explain this result with the “Markedness Differential Hypothesis” (Eckman 1977; as in Rasier & Hiligsmann 2007 discussed in Section 2.1.6), whereby marked structures are more difficult to learn than unmarked ones, and also mention the “Similar Differential Rate Hypothesis” (Major & Kim 1996), according to which marked structures are also acquired with a lower speed of learning. This means that Germans can take advantage of possessing both strategies of accentuation in their native language, i.e. the phonological and the pragmatic one. In contrast, Italian learners of German have a harder task: they have to learn that the distribution of prominences is not necessarily phonological in German and is instead associated with the pragmatic status of the referent. Therefore, the most difficult challenge will be to learn to deaccent the post-focal given referent in nuclear position.

Despite enriching our knowledge of prosodic marking in Italian and German, the studies reviewed above have certain limitations. They investigate a relatively small group of participants, making generalisation of the results difficult, in particular in light of individual differences. They also use a game in which noun phrases are elicited from alternating speakers, where the information status of adjectives and nouns differs across turns. The statements in alternating turns may not have created an engaging interaction between speakers, who may not have assumed the other player’s sentences as the context for their own productions, and, instead, speakers may have concentrated on their own list of statements. Finally, the authors focus on the categorical presence or absence of pitch accents and pitch accent type, and only provide limited information on continuous measures – Swerts et al. (2002) report the F0 excursion only. However, previous research on other languages has shown that a closer inspection of continuous phonetic parameters can provide essential information about the expression of pragmatic contrasts, raising the question as to whether a closer examination of these parameters might have revealed differences that were not captured in the categorical analysis.

In the following study, I attempt to overcome some of these limitations using a semi-spontaneous interactive board game to elicit different types of information structure in a more naturalistic interaction and with a larger sample of participants compared to previous studies. Moreover, I analyse the way speakers modulate continuous parameters to prosodically mark information status and propose a categorical interpretation of the results.

2.2 Method

2.2.1 Data elicitation

As discussed above, some previous studies on the prosodic marking of information status by Italian learners of German elicited data using a card game structured in form of statements between two participants. In that game, both participants receive an equal set of cards containing pictures of different types and varying colours. In alternating turns, one participant picks a card and names its content so that the other participant can align the corresponding card on a board. The two participants alternate the roles of *instruction giver* and *instruction follower*. The variation of picture type (noun) and colour (adjective) was designed to create contrastive information statuses in two successive noun phrases (NPs) produced by participants. However, this type of task presents some disadvantages: it does not favour interaction, so that participants might not assume the other's turn as a context for their own statement and the production of alternating statements might become repetitive and create a list effect. These disadvantages might affect the prosody of participants' realisations and interfere with the pragmatic conditions intended by the experimenters.

Therefore, in the present study special care has been taken in designing an elicitation game to 1) increase the degree of interaction and 2) avoid the risk of repetitiveness. Despite the difficulty in collecting such specific items in more spontaneous conversation, the design was oriented towards the best compromise possible between ecological validity and the elicitation of noun phrases under the intended pragmatic conditions.

To do so, I created a semi-spontaneous interactive board game to be played in pairs. Each participant received a differently randomised board containing 62 sequentially numbered squares. Each square had a flap which could be lifted to see what is underneath, that is images of various types (noun) and colours (adjective). One example board without the flaps is shown in Figure 2.1. All possibly occurring types and colours were listed in the instructions of the game (Figure 2.2 and Figure 2.3) both in visual and written form, but the boards only contained the pictures to avoid interference from reading. Participants were also provided with an additional empty board, displaying only the numbered squares. The task was intended as a distraction, with the aim that participants would pay less attention to their speech. In the instructions, participants were informed that they would go through the table in sequentially alternating turns and that two items are important to win the game, golden apples and bombs. The latter destroys the opponent's golden apples. The person who finds the most golden apples at the

end of the game wins, provided that they have correctly transcribed the content of the other player's board.

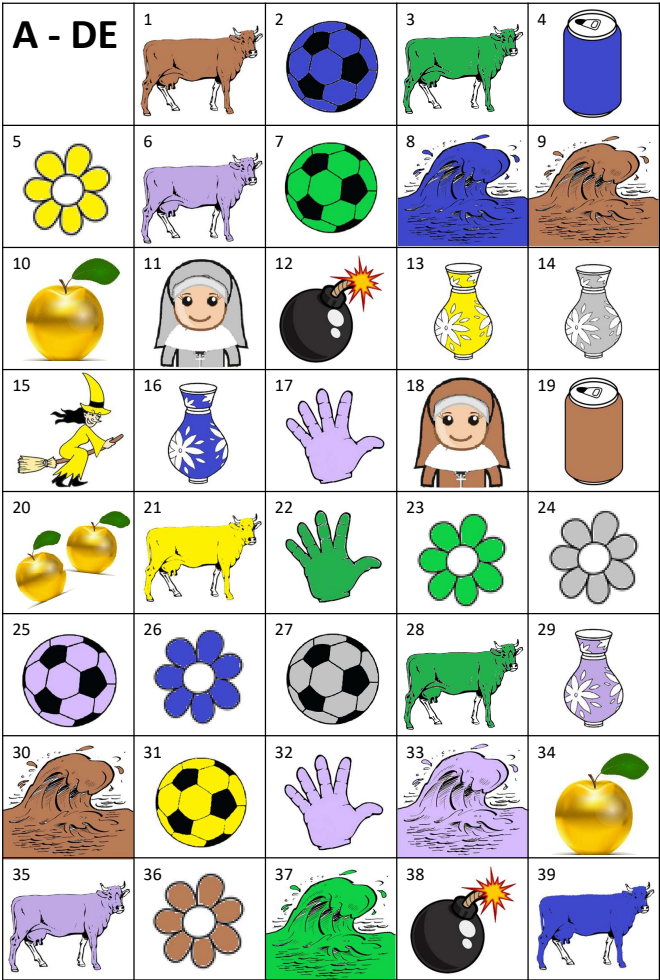


Figure 2.1: Example board of the elicitation game, version for player A in German language. For the Italian example board version of the board game consult Sbranna et al. (2023).

Participants were instructed to sequentially lift the flaps of the table in alternating turns and communicate with a suggested script exemplified in the instructions. The game proceeds as follows: Player A starts uncovering the first picture

by lifting the corresponding numbered flap (square number one) and asks player B if they have the picture they see, using a question in which they mention the type and colour of the image (e.g. “Do you have a yellow cow?”). To answer the question, Player B also uncovers the picture in square number one on their own board, and answers with yes or no, followed by the mentioning of the type and colour of the matching or mismatching image (e.g. “Yes, I have a [matching image]” or “No, I have a [mismatching image]”). This exchange constitutes a game turn. Player B would take the next turn and ask a question about the next square. This alternation of exchanges between players continues until all the squares are revealed. At the end of each turn, both players write on the board (with 62 empty spaces), what their opponent has on their board. At the end of the game, each participant counts how many golden apples they found that were not destroyed by the opponent’s bomb. The person with the most golden apples checks the correctness of their list (the empty board they filled up throughout the game) together with the opponent and wins or loses the game accordingly. This is done in order to keep participants alert and engaged in the game throughout.

Players’ answers about mismatching images by either type or colour, or both type and colour contain our target noun phrases with contrastive elements. The game also elicits yes-replies about matching images by both colour and type. These were inserted only to avoid the bias of a negative answer. Example 1 provides one exchange for each of the possible information structure conditions in the Italian and German versions, with the respective English translation. The questions serve as pragmatic context and the replies as carrier sentences of the noun phrases, marked in bold:

- (1) a. Italian

A: *Hai una **mano nera***?
 A: Do you have a black hand?²
 B: *No, ho una **mano lilla***.
 B: No, I have a lilac hand.
- b. German

A: *Hast du eine **braune Blume***?
 A: Do you have a brown flower?
 B: *Nein, ich habe eine **braune Nonne***.
 B: No, I have a brown nun.

²Note that in Italian the order of the elements in the NP is reversed, i.e. the noun is followed by the adjective, contrary to the English translation.

The pseudo-randomisation of the sequence of images in the boards followed the criterion that no two identical nouns or adjectives could occur at two subsequent turns, in order to avoid an unintended degree of activation of the elements deriving from the preceding turn, interfering with the desired prosodic realisations within the turn.

One final clarification about the information structures of the target noun phrases is necessary. This study is concerned with the dimension of cognitive states (new and given, with all elements being equally accessible since they are listed in the instructions) although, given the context of elicitation (see Example 1), in our items this dimension overlaps with pragmatic functions, i.e. NN with broad focus, and the new element of GN and NG with contrastive focus. The latter case is defined as a contrastive and not a corrective focus to distinguish it from occasional spontaneous occurrences of explicit corrections in our corpus, e.g. instances elicited by a context question like “Have you said red apple?” (with “red” being in focus). Still, I consider contrastive and corrective focus as different degrees of contrast on a continuum and not as two distinct categories. It has also been claimed that it is unlikely that languages prosodically distinguish contrastive and corrective foci. Instead, speakers would increase prominence on contextually salient foci, which may be explicit contrasts or corrections (Baumann et al. 2006; Calhoun 2009; Féry 2013; Kügler & Calhoun 2020).

With this method, I elicited data from Italian and German native speakers (control groups) and Italian learners of German in both their L1 and L2 (Section 1.3.1 and Section 1.3.2 for details on participants and recording sessions, respectively).

2.2.2 Corpus

The target noun phrases were derived from turns in the elicitation game. All NPs are composed of a disyllabic noun and a disyllabic adjective, both with penultimate stress, and correspond to three different types of information structures: given-new (GN), new-new (NN) and new-given (NG). Notice that in Italian the noun precedes the adjective, while in German the adjective precedes the noun. All target NPs are listed by information structure condition and language in Table 2.1. Each speaker uttered each noun phrase only once.

I do not include in the analysis the functional elements which were merely intended to give the game a goal and avoid repetitiveness, i.e. the mentions of golden apples, bombs and noun phrases inserted in yes-reply carrier sentences, whereby both the noun and the adjective are contextually given. The reason for not including this latter given-given condition (GG) is that due to the semi-spontaneous nature of the game, many speakers simply answered “yes” to the

2 Prosodic marking of information status

ISTRUZIONI

Ognuno di voi ha una tabella con 62 finestrelle numerate. Ogni finestrella, se sollevata, mostra un'immagine. La sequenza delle immagini è diversa per ciascun giocatore e le immagini si differenziano per forme e colori.

Due immagini sono particolarmente importanti per vincere: mele d'oro e bombe. Il giocatore che trova più mele d'oro vince, ma ad un'unica condizione, ossia che abbia trascritto la sequenza delle immagini dell'avversario correttamente. Attenzione alle bombe! Se trovi una bomba, puoi annullare una mela d'oro dell'avversario.

Per ricordare quante mele d'oro hai, ogni volta che ne trovi una prendi una carta con la mela d'oro. Quando il tuo avversario trova una bomba, rimettila a posto. In questo modo dovrai semplicemente contare quante carte hai alla fine del gioco.

Per raggiungere l'obiettivo dovete comunicare con l'avversario come nell'esempio seguente:

Apri la finestra n.1, vedi una matita verde e chiedi:
Giocatore A: "Hai una matita verde?"
L'altro giocatore risponde con sì o no e comunica la figura che vede.
Giocatore B: "Sì, ho una matita verde" oppure "No, ho una moneta gialla"

Alla fine del turno entrambi saprete cosa c'è nella finestrella n.1 dell'avversario. Scrivetelo nella tabella vuota a vostra disposizione.

Nel turno successivo chiede per primo colui che ha appena risposto, sempre seguendo l'esempio mostrato. Ciò significa che il giocatore A chiederà sempre per primo con i numeri dispari, mentre il giocatore B chiederà per primo con i numeri pari (mi raccomando non invertitevi!).

Alla fine potrete contare le mele d'oro che avete trovato. Chi ne ha di più controlla di aver scritto correttamente la lista delle immagini dell'avversario. Solo in questo caso potrà vincere, altrimenti ha vinto automaticamente l'avversario!

Buon divertimento e grazie per la vostra partecipazione.

LISTA DEI POSSIBILI OGGETTI E COLORI:

















								
Lilla	Bordeaux	Fucsia	Beige	Blu	Verde	Nero		
								
Nave	Vela	Pera	Luna	Rana	Pane	Pala	Mano	Mela

Figure 2.2: Italian board game instructions with list of all occurring colours and objects.

ANLEITUNG

Jeder Spieler bekommt ein Spielbrett mit 62 nummerierten Fenstern. Jedes Fenster kann geöffnet werden und zeigt ein Bild. Die Reihenfolge der Bilder ist anders für jeden Spieler und die Bilder unterscheiden sich in Form und Farbe.

Zwei Arten von Bildern sind besonders wichtig, um zu gewinnen: goldene Äpfel und Bomben. Der Spieler, der mehr goldene Äpfel findet, gewinnt – unter einer Voraussetzung: er hat die Reihenfolge der Bilder des Mitspielers korrekt aufgeschrieben. Aber Achtung vor den Bomben! Mit einer Bombe kann man einen goldenen Apfel des Mitspielers zerstören und damit ungültig machen.

Um zu wissen, wie viele goldene Äpfel du hast, nimm dir ein Bild mit einem goldenen Apfel jedes Mal, wenn du einen findest. Wenn dein Mitspieler aber eine Bombe hat, leg das Bild zurück. So kannst du einfach die Bilder am Ende des Spieles durchzählen, um zu wissen, wer gewonnen hat.

Um das Ziel zu erreichen, muss man dem Mitspieler die genaue Reihenfolge der Bilder mitteilen, wie im folgenden Beispiel:

Du öffnest Fenster Nr. 1, siehst einen schwarzen Mond und fragst:
Spieler A: „Hast du einen schwarzen Mond?“
Der andere Spieler antwortet mit ja oder mit nein und sagt was er/sie in Fenster Nr. 1 hat:
Spieler B: „Ja, ich habe einen schwarzen Mond“ oder „Nein, ich habe eine grüne Flagge“

Am Ende dieses Zuges wisst ihr beide, was in der Tabelle des jeweils anderen steht und tragt es in eure leere Tabelle ein.

Im nächsten Zug stellt derjenige, der zuletzt geantwortet hat eine neue Frage, wie im Beispiel oben, d.h. Spieler A fragt immer bei ungeraden Nummern, Spieler B immer bei geraden Nummern.

Am Ende könnt ihr die Äpfel zählen. Wer am meisten davon hat, kontrolliert, ob er die Reihenfolge der Bilder in der Tabelle des anderen korrekt aufgeschrieben hat. Nur in diesem Fall, gewinnt er, sonst hat der andere automatisch gewonnen!

Viel Spaß und danke fürs Mitmachen.

LISTE DER MÖGLICHEN OBJEKTEN UND FARBEN:

















							
Grau	Lila	Blau	Braun	Grün	Gelb	Hand	Kuh
							
Welle	Blume	Vase	Nonne	Birne	Dose	Ball	Hexe

Figure 2.3: German board game instructions with list of all occurring colours and objects.

Table 2.1: Target noun phrases by condition for Italian and German.

Condition	Italian	German
GN	mano lilla	braune Welle
GN	nave nera	blaue Blume
GN	mela verde	graue Vase
GN	rana lilla	braune Nonne
GN	vela nera	blaue Birne
GN	luna verde	graue Dose
NN	nave lilla	blaue Welle
NN	mela nera	graue Blume
NN	mano verde	braune Vase
NN	vela lilla	blaue Nonne
NN	luna nera	graue Birne
NN	rana verde	braune Dose
NG	mela lilla	graue Welle
NG	mano nera	braune Blume
NG	nave verde	blaue Vase
NG	luna lilla	graue Nonne
NG	rana nera	braune Birne
NG	vela verde	blaue Dose

context question without following the suggested script and using the noun phrase, in other words, I consider these productions driven by the script only. The first two exchanges were used as a training phase and, therefore, also excluded from the analysis.

It is important to mention that in some cases, speakers were so engaged in the game that they forgot the suggested script for the interaction. For this reason, some tokens were not realised as prescribed, resulting in a few missing items. However, these cases demonstrate that the game succeeded in engaging the speakers, resulting in more spontaneous behaviour than generally expected from a scripted task. Table 2.2 contains the number of items for each condition and language group. For the sake of ecological validity, I decided not to exclude any items from the acoustic analysis based on subjective impression of what a “good” or “bad” item is.

Table 2.2: Amount of noun phrases collected by group and condition.

(Control)	IT L1 (Learners)	IT L1 (Learners)	GE L2 (Target)	GE L1
Total items	231	670	718	323
GN items	69	222	240	108
NN items	81	220	239	108
NG items	81	228	239	107

2.2.3 Measurements

This study makes use of a different approach to the analysis than previously done. The continuous modulation of prosodic parameters – as compared to previous studies with a categorical approach – are analysed by using the open-source ProPer workflow (Albert et al. 2020). This innovative method is based on two corresponding and interacting acoustic time series: F0, which measures the acoustic rate of oscillation of the perceived pitch, and periodic energy, which measures the acoustic strength of the pitch-bearing portions of the signal.

The ProPer workflow derives the periodic energy curve from Praat’s signal processing objects (Boersma & Weenink 2021), which is used in R (R Core Team 2021) to 1) produce periograms, enriched visual representations of F0 trajectories modulated by periodic energy (Albert et al. 2018), and 2) calculate various metrics to account for the F0 movement and the prosodic strength of syllabic intervals. I employ periograms for visualisation and three ProPer metrics to quantify aspects of prosody and perform statistical inference: a) periodic energy mass, b) synchrony and c) $\Delta F0$. I will now explain these metrics in a way functional to the interpretation of the results. Please refer to Sbranna et al. (2023) and Sbranna, Albert, et al. (2025) for technical details on their calculation.

Periodic energy correlates with sonority so that the fluctuations in the periodic energy curve are distributed around sonority peaks, i.e. syllable nuclei (see Albert et al. 2018 for details and visual examples), tending to correspond in this way to syllabic intervals. The prosodic strength of each syllable is given by the periodic energy mass, from here on referred to as “mass”, reflected by the area under the periodic energy curve. Mass accounts for both duration and power of each syllable and is normalised relative to the other syllables within a given utterance, such that values above one indicate strong mass and values below one indicate weak mass (Figure 2.4a).

2 Prosodic marking of information status

Two measurements account for the shape of F0 by calculating interactions between the periodic energy and F0 curves (Cangemi et al. 2019): synchrony and $\Delta F0$. Synchrony reflects the shape of the F0 contour within syllabic units (Figure 2.4b – akin to peak alignment, e.g. Arvaniti et al. 2006), while $\Delta F0$ reflects the shape of the F0 contour with respect to the previous syllabic unit (Figure 2.4c). For both metrics, positive values indicate a rising F0, while negative values signify a falling F0. To normalise these metrics, their relative values are used. Synchrony (measured in milliseconds) is calculated relative to the duration of the containing syllable to accurately represent the F0 slope in syllables of different lengths. $\Delta F0$ (measured in Hertz) is calculated relative to the speaker’s range to minimise speaker-specific paralinguistic effects, such as gender differences. Note that in our analysis, the value of $\Delta F0$ on the first syllable does not refer to the difference from the previous portion of the utterance, which is not taken into account, but rather the difference from the median F0 of that speaker. This value is useful to flag cases starting with a relatively high F0, resulting in a positive $\Delta F0$ value.³

This workflow is applied to the acoustic analysis of the target noun phrases under different information structure conditions.

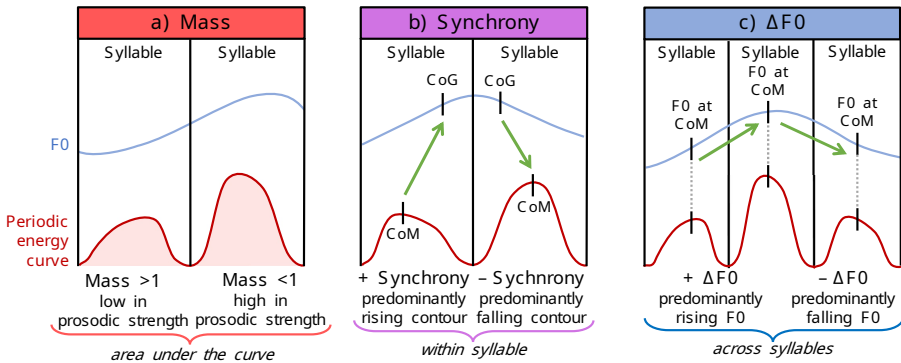


Figure 2.4: Integrated measures of F0 and periodic energy.

³I use this method on the first syllable of the target NPs instead of $\Delta F0$ referring to the previous portion of the utterance because our data often display a pause between the carrier sentence (“No, I have a [...]”) and the target item, i.e. the noun phrase. Since there is not always analysable material preceding the target item, this choice allows me to present a unified measurement for $\Delta F0$ values in Syllable 1 across all data.

2.2.4 Bayesian analysis

Statistical inference was performed by fitting Bayesian hierarchical linear models using the Stan modelling language (Carpenter et al. 2017) and the R package *brms* (Bürkner 2016). For each language group, the differences among conditions in synchrony, $\Delta F0$ and mass were tested as a function of factors *CONDITION* (reference level *NG*), *SYLLABLE* (reference level *SYLLABLE 1*) and their interaction. As random effects, the models include random intercepts for *TOKEN* and *SPEAKER*. For *SPEAKER* the models also include by-speaker random slopes for *CONDITION* and *SYLLABLE* and correlation terms between all random effect components. For models testing the differences across groups, the fixed effect *GROUP* was added to *CONDITION* and *SYLLABLE*, as well as a three-way interaction between them.

For the measurements of synchrony, $\Delta F0$, and mass, a normal distribution was used, and regularising priors for the intercept and the regression coefficient were defined based on theoretical reasons and observations on other datasets. Priors for synchrony are theoretically driven: The two centres, CoM and CoG, are both attracted to the centre of the interval, so the distance between them does not tend to exceed 25% of the entire duration of the interval containing them. Priors for $\Delta F0$ are based on observations in multiple data sets, in which most of the $\Delta F0$ values reflecting the $F0$ change between syllables are below 50% (of that speaker's range). Higher values are possible but mostly do not exceed 70%. Priors for mass are also based on observations in multiple data sets, in which the vast majority of values is found between 0.25–2.5.

The default settings of the *brms* package were retained for all other parameters. For relative synchrony, the intercept was set at $\mu = 0$, $\delta = 30$ and the regression coefficient at $\mu = 0$, $\delta = 5$; for relative $\Delta F0$, the intercept was set at $\mu = 0$, $\delta = 50$ and the regression coefficient at $\mu = 0$, $\delta = 25$, for mass, the intercept was set at $\mu = 0$, $\delta = 3$ and the regression coefficient at $\mu = 0$, $\delta = 0.5$. Four sampling chains for 4000 iterations with a warm-up period of 3000 iterations were run for all models. There was no indication of convergence issues (no divergent transitions after warm-up; all $Rhat = 1.0$), including from visual inspection of the posterior distributions.⁴

For all relevant contrasts (δ), I report the expected values under the posterior distribution and their 95% credible intervals (CIs), i.e. the range within which an effect is expected to fall with a probability of 95%. For the difference between each contrast, the posterior probability that a difference is bigger than zero ($\delta >$

⁴The model's assumption for $\Delta F0$ were not fully satisfied since our posterior simulations are less leptokurtotic than our actual data. Still, the model does not show convergence problems.

0) is also reported to ensure comparability with conventional null-hypothesis significance testing. In particular, it is assumed that there is (compelling) evidence for a hypothesis that states $\delta > 0$ if zero is (by a reasonably clear margin) not included in the 95% CI of δ and the posterior $P(\delta > 0)$ is close to one (cf. Franke & Roettger 2019).

All models, results and posteriors can be inspected in the accompanying RMarkdown file at the Open Science Framework (OSF) repository (<https://osf.io/9ca6m/>).

2.3 Results: Italian L1

2.3.1 Learners of German

Figure 2.5 shows three representative example periograms (Section 2.2.3), along with the three acoustic metrics, one for each information structure condition as uttered by Italian learners of German in their L1 Italian. By visualising periograms, it can be observed that GN and NN are similar in that F0 is rising throughout the first syllable and reaches a peak on the second syllable, after which there is a fall. By contrast, NG reaches a peak already in the first syllable, which is where it starts falling. Thus, there are two intonation patterns that are distinguished through timing of the F0 fall: earlier when the final position features given information (NG) and later when the final position features new information (GN and NN). This distinction carries over to the transition between words, i.e. between the second and third syllables. The falling F0 trend ends earlier in NG, such that the transition from the second to the third syllable is mostly flat, while it is still falling in GN and NN. As for the values of mass, the syllables associated with stress (Syll1 and Syll3) tend to be stronger than the following unstressed syllables (Syll2 and Syll4, respectively), and the stressed syllables in new words of the NG and GN conditions tend to display stronger energy than the stressed syllables in adjacent given words.

Aggregated data of synchrony, $\Delta F0$ and mass confirm these observations (Figure 2.6). Specifically:

- *Synchrony*. Positive synchrony values indicating a rising F0 trajectory can be found almost exclusively on Syll1 of GN and NN. The earlier F0 fall in NG is reflected in negative synchrony values on Syll1 and Syll2, while the later F0 fall in GN and NN is reflected in negative synchrony values on Syll2 and Syll3.

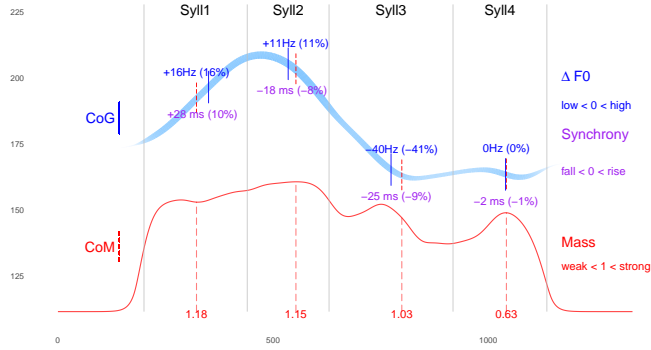
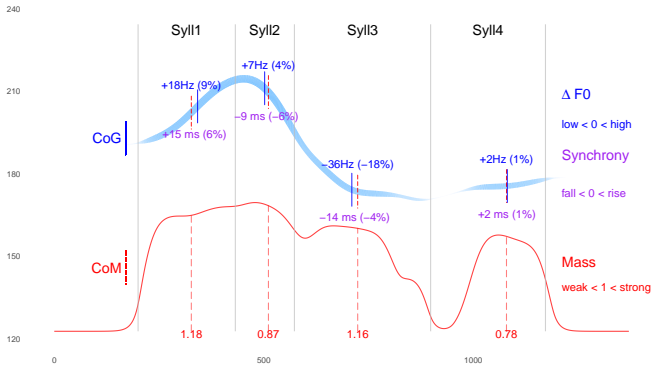
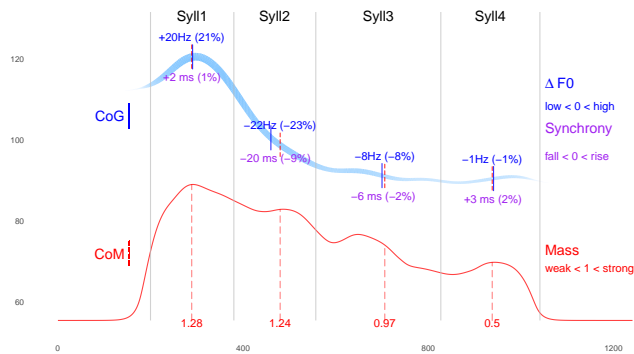
(a) GN. The token is *mano lilla* ('lilac hand').(b) NN. The token is *rana verde* ('green frog').(c) NG. The token is *luna lilla* ('lilac moon').

Figure 2.5: Example perigrams displaying F0 and periodic energy for two-word noun phrases in three information structure conditions produced by L1 Italian learners.

2 Prosodic marking of information status

- $\Delta F0$. Positive $\Delta F0$ values on Syll2 and negative $\Delta F0$ values on Syll3 are indicative of the location of the F0 peak in Syll2 of GN and NN conditions. In contrast, negative $\Delta F0$ values on Syll2 in the NG condition indicate that the F0 peak is within Syll1.
- *Mass*. Mass values generally reflect that stressed syllables and new words promote stronger prosodic strength, as expected. Stressed syllables are reflected in the distinction between Syll1 (stressed) vs. Syll2 (unstressed) as well as Syll3 (stressed) vs. Syll4 (unstressed), while information status (given vs. new) is mostly reflected in the distinction between Syll1 vs. Syll3 of the GN and NG conditions. These trends are clearly apparent in the NG condition and, to a lesser extent, in the NN condition. Mass values in the initial three syllables of the GN condition display a wide distribution of mostly strong energy (values above one), thus appearing to attenuate the stress-related mass distinctions in the first word and the information status mass distinctions between the two stressed syllables (Syll1 vs. Syll3).

These trends revealed to be robust by means of Bayesian analyses:

- *Synchrony* on Syll1 presents lower values in NG as compared to GN ($\delta = 5.1$, CI [4.13; 6.07], $P(\delta > 0) = 1$) and NN ($\delta = 6$, CI [5.01; 6.94], $P(\delta > 0) = 1$), while synchrony on Syll3 presents higher values in NG as compared to GN ($\delta = 3.46$, CI [2.49; 4.42], $P(\delta > 0) = 1$) and NN ($\delta = 4.3$, CI [3.26; 5.23], $P(\delta > 0) = 1$).
- $\Delta F0$ on Syll2 is lower in NG as compared to GN ($\delta = 21.73$, CI [17.64; 25.97], $P(\delta > 0) = 1$) and NN ($\delta = 24.05$, CI [19.93; 28.27], $P(\delta > 0) = 1$), while $\Delta F0$ in Syll3 is higher in NG as compared to GN ($\delta = 15.46$, CI [11.42; 19.37], $P(\delta > 0) = 1$) and NN ($\delta = 19.86$, CI [15.71; 23.74], $P(\delta > 0) = 1$).
- *Mass* on Syll1 is higher in the NG condition than in GN ($\delta = 0.21$, CI [0.17; 0.25], $P(\delta > 0) = 1$) and NN ($\delta = 0.15$, CI [0.11; 0.18], $P(\delta > 0) = 1$). In addition, within NG values of mass are higher in Syll1 than in Syll3 ($\delta = 0.19$, CI [0.13; 0.25], $P(\delta > 0) = 1$).

Both the data and the models strongly support our claim that (Neapolitan) Italian prosodically marks post-focal given elements within NPs through the continuous modulation of acoustic parameters. In particular, NG is realised with an F0 peak early in the first word as opposed to GN and NN, which are realised with an F0 peak late in the first word. These F0 shapes are accompanied by distinct

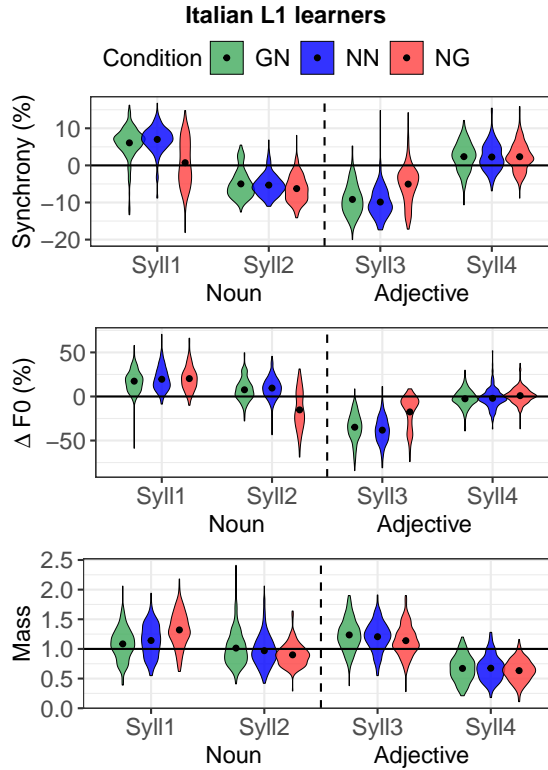


Figure 2.6: Aggregated values of synchrony, $\Delta F0$ and mass (on the y-axes) pooled across Italian L1 learners. The x-axis displays the four syllables of the noun phrases, with Syll1 and Syll2 being the noun and Syll3 and Syll4 the adjective. Information structure conditions are colour-coded: green for given-new (GN), blue for new-new (NN) and red for new-given (NG).

modulations of prosodic strength: along with the earlier F0 peak, NG displays stronger energy on the first syllable as compared to GN and NN. However, modulations are slight and do not result in overall different patterns across information structure conditions. As a consequence, no reduction in energy on the post-focal second word is found as typical of West-Germanic languages.

2.3.2 Monolingual control group

Since the Italian speakers commented above are learners of L2 German and recordings were performed at the Goethe Institute, which is a German-language-dominant setting, a group of monolingual Italian native speakers was recorded

as well to control for a possible influence of the German language and setting on learners' L1 Italian productions. By "monolingual" I mean that these Italian participants, despite some inevitable contact with foreign languages in the past years of their lives (such as in school contexts), were neither learners of, nor proficient or fluent in any foreign language and were not familiar with the Institute. To highlight the fact that there is no exposure to an L2 which could influence their native Italian prosody and that these participants are a different group of L1 Italian speakers than the one presented above (who are learners of L2 German), I refer to this group as the "monolingual control group".

Figure 2.7 shows the same patterns of synchrony, ΔF_0 and mass observed in the Italian learners' dataset for the monolingual control group. Two F_0 shapes distinguished by the location of the peak characterise the realisation of NG vs. GN and NN (earlier vs later in the first word). The stress pattern does not show differences from the learners' group either and reflects the two expected effects in which stressed syllables and new words promote stronger prosodic energy. Information status contrasts are particularly apparent in NG with a proportional reduction in strength on the given element (Syll3) as compared to the new one (Syll1). Akin to the results for the learners' group, this contrast is less clear in GN, with strong mass values across the first three syllables, reducing both stress and information status contrasts. Overall, there were no major differences between the Italian monolingual control group and the Italian learner group.

The similarity observed between the two native Italian groups (learners vs. monolinguals) was confirmed using Bayesian models for each acoustic parameter, suggesting no robust difference across groups overall. The only between-group difference found in a position relevant for the two F_0 contours concerns ΔF_0 , which has higher values in the NG condition on Syll1 and lower values on Syll2, compared to the learners' group. This shows that the control group uses a wider range of F_0 values for the early peak contour, starting higher on Syll1 and reaching lower values on Syll2, which, in turn, is indicative of a steeper slope. However, this variability across groups might not be directly linked to the linguistic background of the speakers. Instead, it could also be ascribed to idiosyncratic differences, given that ΔF_0 values are normalised on each speaker's range and not across speakers.

2.3.3 Summary

To summarise, both the data and the models strongly support our claim that (Neapolitan) Italian does prosodically mark post-focal given elements within noun phrases through the continuous modulation of acoustic parameters, which

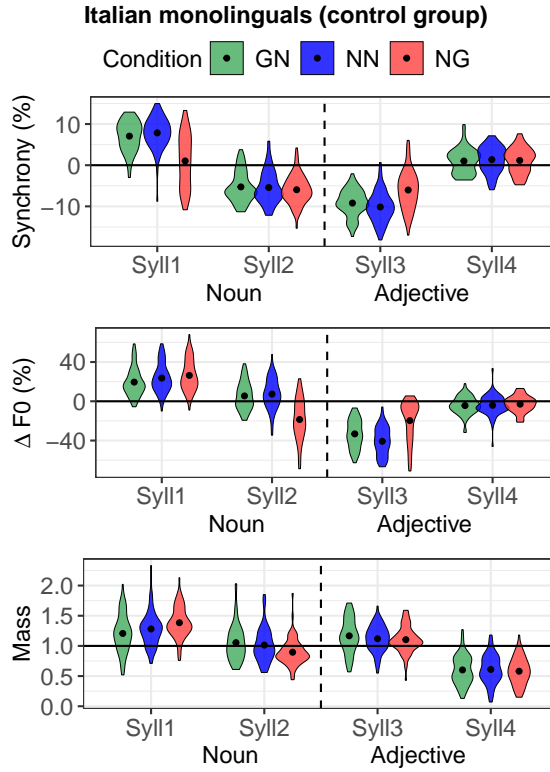


Figure 2.7: Aggregated values of synchrony, $\Delta F0$ and mass (on the y-axes) pooled across Italian monolinguals (control group). The x-axis displays the four syllables of the noun phrases, with Syll1 and Syll2 being the noun and Syll3 and Syll4 the adjective. Information structure conditions are colour-coded: green for given-new (GN), blue for new-new (NN) and red for new-given (NG).

did not emerge in previous categorical analyses of Italian. In particular, NG is realised with a different F0 shape as opposed to GN and NN, i.e. with an F0 peak early (vs. late) in the first word. These F0 shapes are accompanied by distinct modulations of prosodic strength: along with the earlier F0 peak, NG displays stronger energy on the first syllable as compared to GN and NN. Moreover, prosodic strength in the NG condition exhibits the expected stress patterns (stressed syllables are stronger than unstressed ones), as well as the expected information status pattern (despite both stressed syllables displaying values in the domain of strong energy, the stressed syllable of the new word is stronger than the stressed syllable of the given word).

The minor differences found across the two groups did not generally result in different intonation or strength patterns. Overall, it can be confidently claimed that the two Italian native groups, i.e. the Italian learners of German and the monolingual Italian control group, produced prosodically very similar utterances. This result reassures us that the German-language dominant setting and the German linguistic knowledge of Italian learners of L2 German did not exert an influence on their native Italian speech. As a consequence, I will take into account all native Italian data, i.e. by both learners and monolinguals, when discussing native Italian speech itself (as in Section 2.4), while I will use the Italian spoken by learners themselves as a baseline for comparing their interlanguage to their native and target languages (in Section 2.7).

2.4 Perceptual validation for Italian L1

The present finding that Italians use two different intonation contours to mark new vs. post-focal given information in the last position of a noun phrase diverges from previous results on Italian claiming that no prosodic marking of information status is realised within NPs in production. This was previously confirmed by a perception experiment in which Italian listeners could not match the NPs to their information structure from the acoustic signal and in absence of any other contextual cues (Krahmer & Swerts 2008). With this background in mind, a similar small-scale perception experiment was conducted to ensure the reliability of the elicited data. The aim is to verify the representativeness of the noun phrases elicited and to explore whether listeners make use of the prosodic distinctions found in production. Afterwards, an acoustic analysis of the items that successfully communicate the pragmatic function, i.e. correctly matched to their information structure, was carried out to see whether the trends identified in production are confirmed in perception.

2.4.1 Procedure

The perception test included 754 output noun phrases of the production experiment in L1 Italian as auditory stimuli. From a total of 901 items, 147 items presenting disfluency phenomena (empty or filled pauses and final lengthening between the two words of the noun phrase) and extra-linguistic events (laughing, coughing, tongue clicks, microphone noises, etc.) were excluded to avoid interference with listeners' judgment. The exclusion criteria listed were strictly followed and,

in line with the choice made for the production experiment, no item was excluded on the basis of the experimenter's subjective impression of a "good" or "bad" item.

The audio files serving as stimuli were normalised at -23LUFS (a standard reference level for loudness) and their sequence was randomised. To avoid a learning effect, no more than two items with the same information structure appeared sequentially. The online questionnaire was generated using SoSci Survey (Leiner 2019) and was made available to users via www.soscisurvey.de. At the beginning of the questionnaire, participants were explicitly requested to use headphones and to keep the volume level constant throughout the experiment.

The participants were three native speakers of Neapolitan Italian (the same variety as the Italian production data), who were trained phoneticians and not familiar with the present study. They were asked to listen to a sequence of elicited noun phrases composed of a noun and an adjective, and to indicate which element they perceived as new. A written example dialogue from the elicitation game for each information structure condition (reproduced here along with its English translation) was included in the instructions to clarify the context in which the items used as stimuli were produced:

Giocatore A: Hai una mela nera?

Player A: Do you have a black apple?

Giocatore B (condizione dato-nuovo): No, ho una mela rossa.

Player B (given-new condition): No, I have a red apple.

Oppure:

or:

Giocatore B (condizione nuovo-nuovo): No, ho una pera rossa.

Player B (new-new condition): No, I have a red pear.

Oppure:

or:

Giocatore B (condizione nuovo-dato): No, ho una pera nera.

Player B (new-given condition): No, I have a black pear.

For perceptual judgment, participants could choose among the following options: 1) noun, 2) adjective, 3) both noun and adjective or option 4) "I do not know". The test was preceded by a training phase with four practice trials, followed by one single session that could be paused and resumed again at any time.

2.4.2 Results

Recall that the analysis of the production data had revealed two distinctive trends across conditions, i.e. an early vs. late F0 peak on the first word for NG vs. GN and NN information structures, respectively. In other words, when the last element of the NP is post-focal and given, it seems to be marked by an alignment of the F0 peak within the first syllable, while the F0 peak is aligned later in the first word when the last element of the NP is new.

The results of the perception experiment are shown in Figure 2.8. The intended information structure in which the stimuli were elicited during the game is on the y-axis, while the ratings given by listeners are on the x-axis. From the percentages yielded by the matrix, it emerges that GN and NN items are frequently confused, whereas agreement is consistently more robust for NG items, reflecting the two F0 patterns found in production. In particular, the matrix shows that 53.8% of GN stimuli were recognised as such and 42.7% were matched to a NN information structure. Likewise, 50.3% of NN stimuli were recognised as such and 45.9% were matched to GN. For both GN and NN, judgements were thus nearly at chance level. This result from perception corresponds to the results of the acoustic analysis showing that GN and NN share a very similar prosodic contour that makes it difficult to reliably associate the stimuli to one of the two pragmatic conditions. Moreover, both GN and NN stimuli were matched to NG only in 1% of cases, meaning that listeners could discriminate the shared contour with a later F0 peak on the first word very well and did not confuse it with NG. NG was correctly recognised in 64.1% of cases and, only matched to GN and NN in a minority of cases (14.4% and 20.8%, respectively). An explanation for these mismatches is offered by the acoustic analysis reported in the following paragraph, showing that the two F0 contours can indeed be easily distinguished in perception.

2.4.3 Acoustic analysis of correctly identified items

In total, 248 items were correctly matched to the intended information structure by all three listeners. Of these, 41 were GN, 36 NN and 171 NG noun phrases, confirming greater success in correctly identifying NG. The acoustic analysis of these items represents a complementary explanation for the results of the perceptual validation test, both for cases of correct and incorrect attribution. In particular, it offers a clearer overview of how far the acoustic parameters are differently modulated in the two F0 patterns to be perceptually discriminated and, in turn, shed light on the acoustic characteristics of the items which were not correctly matched to information structure.

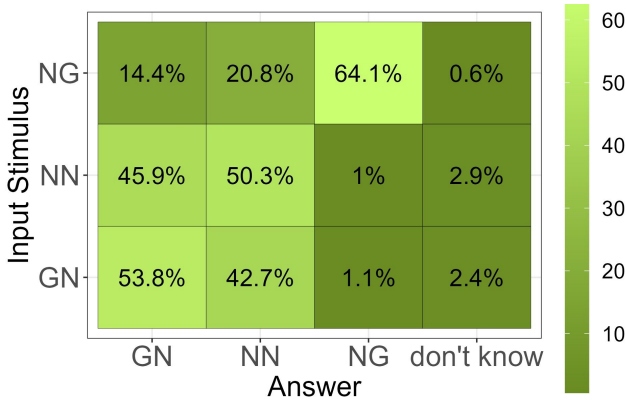


Figure 2.8: Ratings from the perception test pooled across listeners. The actual information structure of stimuli elicited in production is displayed on the y-axis. The information structure matched to the stimuli by listeners is displayed on the x-axis.

Figure 2.9 displays the values of synchrony and ΔF_0 for the stressed syllable of the last element in the NP (Syllable 3) of the correctly identified items and, for comparison, of all production data. Syll3 was chosen because both the distributions of synchrony and ΔF_0 show the effects of the different F_0 peak alignments across conditions. Specifically, the F_0 fall is already completed before Syll3 in the case of the early peak (NG), meaning that synchrony and ΔF_0 values should mainly be distributed around zero indicating no F_0 movement at this location. In contrast, the F_0 fall takes place on Syll3 in the case of the late peak (GN and NN) so that synchrony and ΔF_0 will be distributed in the domain of negative values, indicating a falling F_0 within Syll3 and across Syll2 and Syll3, respectively. Indeed, this is the pattern found after filtering out items that were not reliably matched in the perception test, visible in the lower panel, where the differences in synchrony and ΔF_0 are strengthened. The more negative values for the NG condition (ΔF_0 values below -20% and synchrony values below -10) yielded by NG items realised with a late F_0 peak and visible in the upper panel almost disappear in the lower panel. This means that those instances of NG which were not realised with an earlier F_0 peak, were also not recognised as conveying NG information structure and, instead, fell under the minority of NG cases matched to GN and NN information structures. I speculate on the reasons for such occurrences in this corpus in the discussion.

The aggregated values of mass were also examined more closely (Figure 2.10). Unlike synchrony and ΔF_0 , for which the trends in the large dataset were confirmed and strengthened in the perception-based reduced dataset, the values of

Synchrony and $\Delta F0$ for Syllable 3

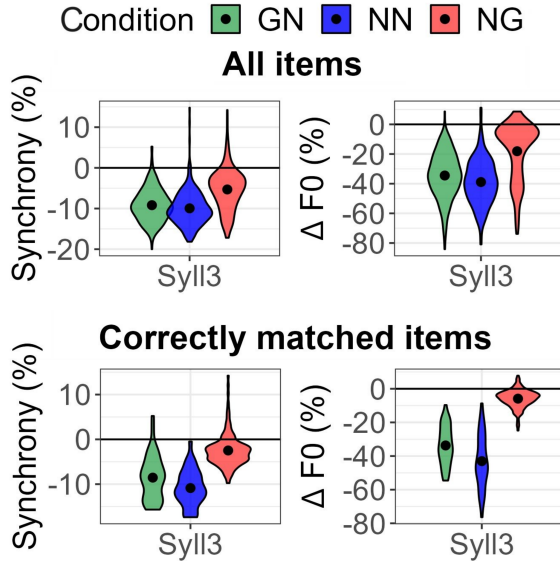


Figure 2.9: Synchrony and $\Delta F0$ values at syllable three for all production data (upper panel) vs. only items that were correctly matched to the intended information status (lower panel). The x-axis displays syllable three, the last lexically accented syllable in the noun phrase. Information structure conditions are colour-coded: green for given-new (GN), blue for new-new (NN) and red for new-given (NG).

mass for the reliably matched items reveal new patterns that were not evident before. A reduction in strength on the given element had already been observed within the NG condition (Syll1 > Syll3), while GN displayed mostly strong energy on both stressed syllables, thus appearing to attenuate information status distinctions between the two words of the noun phrase (visible in the upper panel). Instead, the perception-based dataset in the lower panel shows a similar attenuation of the given element also within the GN condition, where the first syllable, even if lexically stressed, presents mostly weak energy and, therefore, contrasts with the stressed syllable of the new word (Syll3) bearing strong mass, in particular.

As a result, the reduced dataset clearly shows, for both NG and GN, a proportional reduction of mass on the given element in comparison to the new element. However, the distribution of values on the given elements across GN and NG

(Syll1 and Syll3, respectively) differs in the extent of their proportional reduction in energy compared to the respective new elements (Syll3/Syll1). In particular, the stressed syllable of the given word in NG (Syll3, corresponding to the nuclear position) still shows a distribution in the domain of strong mass values, while values of the stressed syllable of the given word in GN (Syll1, corresponding to a pre-nuclear position) are mostly distributed in the domain of weak mass values. This difference can also be seen in their mean values, represented by the black dots on the violins: below one on Syll1 in GN and above one on Syll3 in NG.

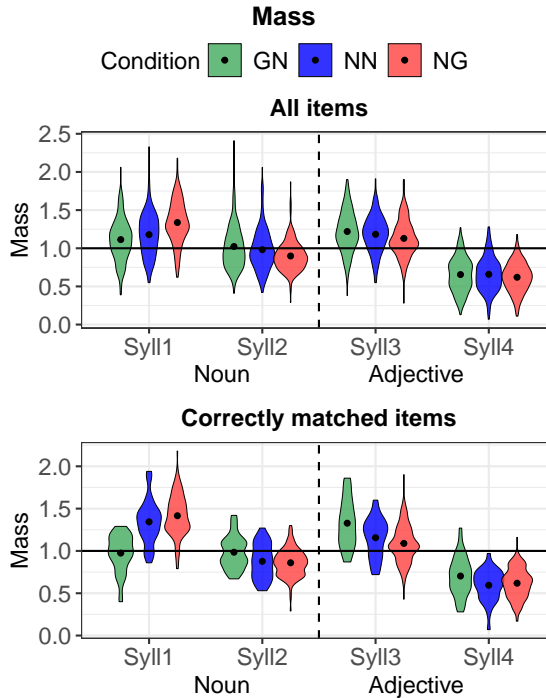


Figure 2.10: Aggregated values of mass for all Italian L1 production data (upper panel) vs. only items that were correctly matched to the intended information status (lower panel). The x-axis displays the four syllables of the noun phrases, with Syll1 and Syll2 being the noun and Syll3 and Syll4 the adjective. Information structure conditions are colour-coded: green for given-new (GN), blue for new-new (NN) and red for new-given (NG).

Differences in mass values for the correctly matched items was further tested by fitting a Bayesian hierarchical linear model on this reduced data set, which

demonstrated that the observed patterns are statistically robust. In the GN condition, the given element (Syll1) is reduced compared to the new element (Syll3) ($\delta = 0.28$, CI [0.17; 0.39], $P(\delta > 0) = 1$). Moreover, at the same location (Syll1), the stressed syllable of the given element in GN is robustly weaker than the stressed syllable of the new element in NN ($\delta = 0.35$, CI [0.28; 0.41], $P(\delta > 0) = 1$).

This finding suggests that speakers may be using prosodic strength when decoding information status. In particular, mass might be especially helpful for discriminating GN from NN. Both conditions present highly similar F0 contours, but the mass distributions on Syll1 show opposite trends, i.e. weak mass when corresponding to given information status (first word in GN) and strong mass when corresponding to new information status (first word in NN).

2.4.4 Discussion

A small-scale perception study was conducted to verify the reliability of the production data. Overall, ratings suggest that listeners do make use of the distributed modulation of prosodic parameters used by speakers to mark post-focal given information and distinguish between different information structure conditions.

In line with the production data, ratings show that two F0 patterns can be identified in perception: a late F0 peak in the first word of the NP for GN and NN conditions, which share this highly similar F0 contour and are often confused with each other, and an early F0 peak on the first syllable for NG, which is correctly identified more often. A further acoustic analysis of the items correctly matched to their information structure has shown that the NG items which were matched to a GN or NN condition were realised with a later F0 peak alignment. A possible reason why these items are found in the corpus could be that marking information status prosodically was not strictly necessary for the successful transmission of the linguistic message in the context of our elicitation game. Therefore, in a minority of cases, speakers did not prosodically mark the new information shifted from the nuclear position in NG and, instead, realised an unmarked prosodic contour, in which new information corresponds to the nuclear position as well, as in GN and NN conditions. With regard to prosodic strength, the acoustic analysis of the correctly matched items confirmed the pattern found in production for NG, and revealed a new pattern for GN. In production, a proportional attenuation according to information status could be observed only for the NG condition, with both stressed syllables being strong, but with the one corresponding to the given word (Syll3) being less strong than the one corresponding to the new word (Syll1). For GN, a similar attenuation of the given element via mass was revealed in the perception-based, reduced dataset: looking at the items

which were correctly matched to GN by all three raters, the GN items correctly identified are those with the lowest energy values on the given element (Syll1), with a distribution mostly shifted towards the domain of weak energy.

This result can be explained by the fact that the NG information structure is marked by an anticipated F0 peak which is apparently very salient and easily recognisable in perception. Therefore, any other cue would be redundant. Differently, GN and NN conditions, which share the new element in last position, but differ for the information status of the first element, are realised with a highly similar F0 contour so that listeners seem to rely on broad-range energy modulations, when present, to distinguish these two information structure categories.

The measure used to operationalise prosodic strength, i.e. periodic energy mass, results from a calculation accounting for duration and intensity. Thus, it can be taken as one cue to the postfocal attenuation of given information, which has previously been interpreted as deaccentuation when entire phrases are postfocal. Interpreting strong energy as a possible cue to accentuation means that these findings are in line with previous studies reporting some occurrences of prefocal deaccentuation in Italian (Avesani et al. 2015). This is explained by the fact that the postfocal given position in NG corresponds to the metrical head of the noun phrase, which has to bear an accent. On the contrary, an accent corresponding to the prefocal given position in GN is optional (as suggested by Avesani et al. 2015; Bocci & Avesani 2011).

Finally, particularly relevant methodological implications need to be mentioned. Despite including noise in the acoustic analysis of our production data, consistent and robust patterns could be identified in production and confirmed in perception. This shows that it is possible to 1) reduce the experimenter's control on the task output to favour a more spontaneous and interactive behaviour in the experimental setting (in our case the semi-spontaneous board game resulted in a few lost items due to some speakers' spontaneous reactions to the events of the game, but allowed us to observe phenomena which were not previously identified); 2) have a more ecologically valid approach without excluding data considered not representative or stereotypical since they are a possible and real outcome; 3) reliably use periodic-energy-based metrics, which were shown to be robust to noise since the trends found in the entire corpus are confirmed and even more plainly manifested for the correctly matched items only. Of course, this being a small-scale experiment with expert listeners as raters, these results have to be taken as preliminary. Since this study is aimed at exploring whether the modulation of continuous prosodic parameters reveals patterns for the marking of postfocal givenness in Italian learners of German, I will not deal with this issue in greater depth within this context. However, the preliminary perceptual

results warrant further investigation as well as large-scale replication with naive listeners to examine to which extent speakers robustly produce and listeners robustly attend to patterns of weak or strong mass on the first syllable of the NP as a strategy for discerning the GN and NN conditions.

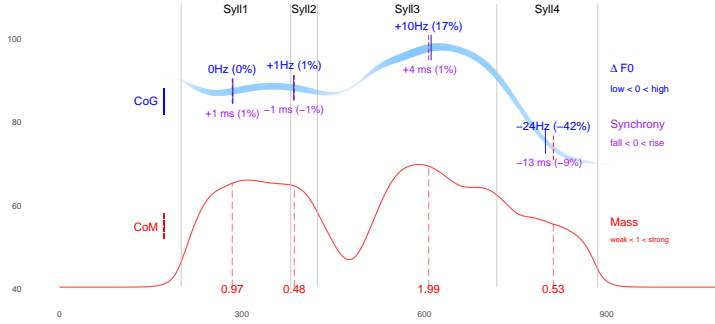
2.4.5 Summary

A production experiment and a perceptual validation test were conducted to explore whether and how Italian speakers (Neapolitan variety) use continuous prosodic parameters to mark information status within noun phrases. The present findings contrast with previous studies based on a categorical analysis of accentuation, in which it was concluded that Italians neither mark information status in production, nor decode it in perception. The analysis presented here, based on periodic-energy-related measures, has brought to light that Italian speakers (both learner and control groups) do mark post-focal given information within noun phrases. However, they do so early in the phrase, modulating the F0 shape on the first word, thus anticipating the upcoming given item. This result is further validated by the results of a perceptual evaluation, showing that NG information structure was the only one matched to prosodic contours presenting an early F0 fall on the stressed syllable of the first element of the NP.

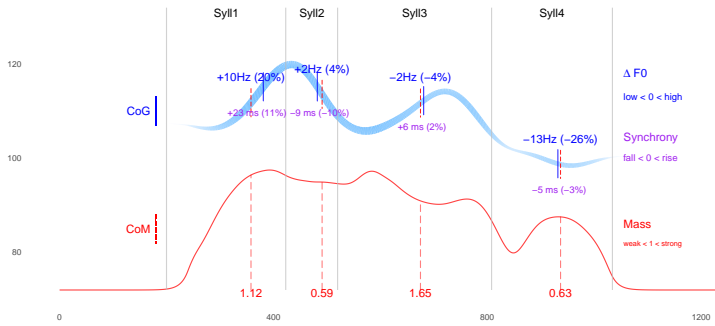
2.5 Results: German L1

Figure 2.11 shows three example periograms (Section 2.2.3), along with the three acoustic metrics, for each of the three information structure conditions uttered by German native speakers, who represent the ideal target of the Italian learners of L2 German. These periograms show F0 and mass patterns which, according to the literature, are typical in this context, with an F0 peak and an increase in energy marking new or focused elements. In the GN and NG conditions, a peaking F0 movement corresponds to the stressed syllable of the new element (Syll3 in GN and Syll1 in NG), together with strong energy. In the NN condition (where both words of the noun phrase are new) two peaking F0 movements – the second lower than the first due to declination – are realised on the two stressed syllables (Syll1 and Syll3) and accompanied by increased energy.

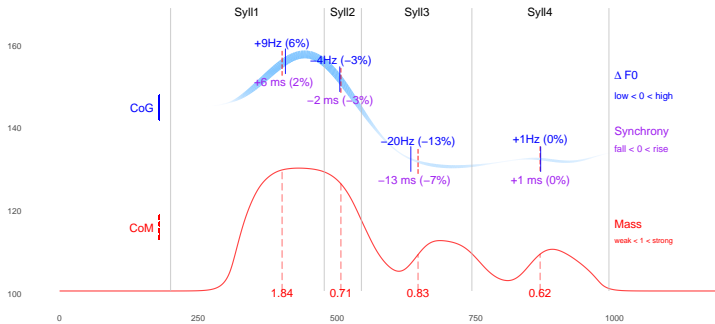
In line with previous findings, these three example utterances show that German native speakers prosodically mark information status by means of an F0 peak and strong energy corresponding to the lexical accented syllable of the new or focused element. Conversely, the lexical accented syllable of the given element is attenuated by means of flat F0 and reduced energy.



(a) GN. The token is *blaue Birne* (Eng. 'blue pear').



(b) NN. The token is *braune Vase* (Eng. 'brown vase').



(c) NG. The token is *graue Nonne* (Eng. 'grey nun').

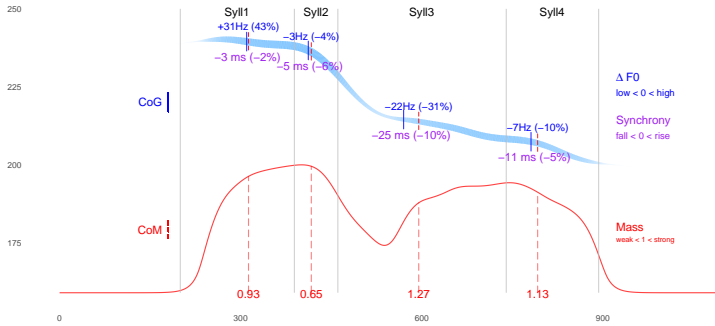
Figure 2.11: Example periograms displaying F0 and periodic energy for two-word noun phrases in three information structure conditions produced by L1 German speakers.

At this point, a side note on variability for the native German corpus is in order (see Figure A3 in the Appendix for the variable L1 German by-speaker contours and compare to the very homogenous L1 Italian by-speaker contours in Figures A1–A2). The example periograms shown in Figure 2.11 display the three most typical realisations described in the literature. However, alternative prosodic realisations were also found in the corpus due to speaker-specific preferences. The highest variability is encountered for GN and NN conditions (see example periograms in Figures 2.12 and 2.13), while the pattern described above for NG (with the F0 peak located on the first word and a flat, low-energy post-focal region) was the most consistent across speakers. In particular, GN and NN can display a hat pattern (Figure 2.13), with high F0 stretching across the two elements of the NP and falling on the third syllable, i.e. after having reached the lexically stressed syllable of the second word.

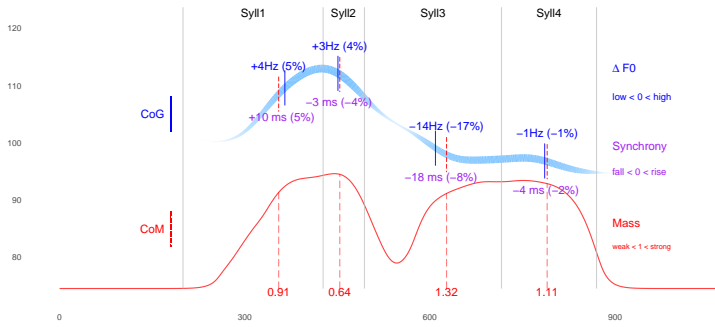
Aggregated values of synchrony, $\Delta F0$ and mass (Figure 2.14) reflect the NG contour displayed in Figure 2.11 and the GN/NN hat pattern shown in Figure 2.13:

- *Synchrony.* In GN and NN, large distributions including negative and positive values on Syll1 and Syll3 reflect the variability described above, i.e. either Syll1 or Syll3 can present an F0 peak. Mean values near zero reflect the hat pattern across the two words until the second lexically stressed syllable, after which clearly negative distributions on Syll4 for both conditions indicate the F0 fall. In NG, mostly negative values are already found on Syll2, showing a falling movement that starts earlier than in GN and NN.
- $\Delta F0$. In GN and NN, negative distributions on Syll4 signal that the F0 fall is realised across Syll3 and Syll4 (before this position, values around zero reflect a somewhat plateauing pattern). In NG the negative distribution on Syll3 signals an earlier location of the F0 fall across the two words, namely between Syll2 and Syll3.
- *Mass.* Mass patterns reflect stress-related distinctions only in the GN and NN conditions, with stressed syllables (Syll1 and Syll3) being strong and unstressed ones (Syll2 and Syll4) weak. In NG, stress-related distinctions on the second word are neutralised by weak mass on Syll3, which, in turn, reflects information status contrasts, as Syll3 is the stressed syllable of the given word. Information-status-related distinctions can be seen to a lesser extent in GN, where both Syll1 and Syll3 present strong mass, but Syll3 is relatively stronger than Syll1. Even if the contrast is not observable for NN

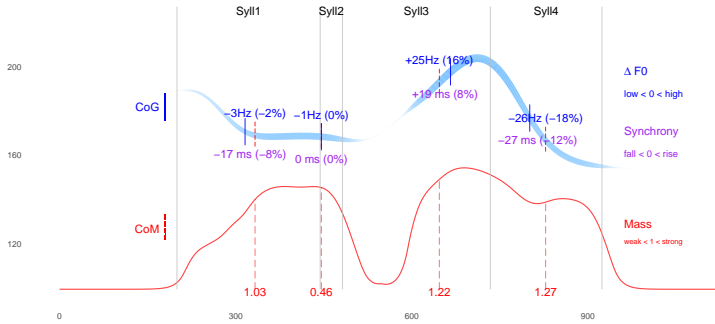
2.5 Results: German L1



(a) GN with falling contour. The token is *blaue Blume* (Eng. 'blue flower').



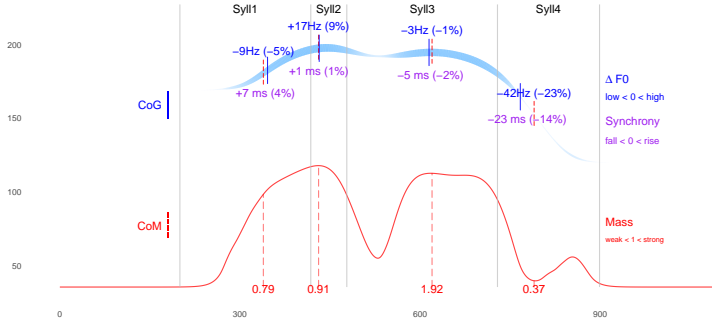
(b) NN with peak on the first word. The token is *graue Birne* (Eng. 'grey pear').



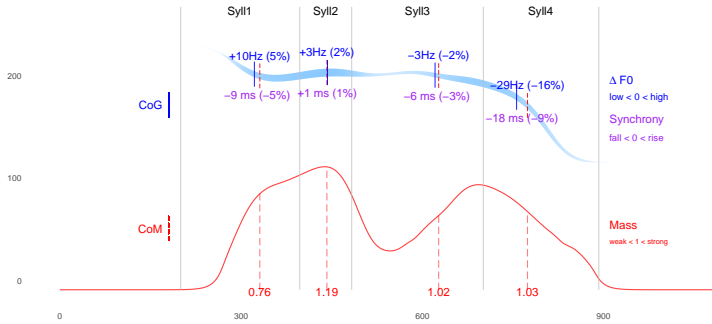
(c) NN with peak on the second word. The token is *graue Blume* (Eng. 'grey flower').

Figure 2.12: Example perigrams in L1 German displaying F0 and periodic energy for alternative contours in GN and NN conditions.

2 Prosodic marking of information status



(a) GN. The token is *graue Dose* (Eng. 'grey can').



(b) NN. The token is *graue Blume* (Eng. 'grey flower').

Figure 2.13: Example periograms in L1 German displaying F0 and periodic energy for the hat pattern in GN and NN conditions.

(with both words having the same information status), it is worth noticing that distributions on both Syll1 and Syll3 are similar independently of their position in the noun phrase (nuclear or pre-nuclear), which might be a further indication that mass is primarily modulated according to information status.

These trends relative to synchrony, $\Delta F0$ and mass were found to be robust by means of Bayesian analyses:

- *Synchrony* on Syll3 is lower in NG than in GN ($\delta = 2.4$, CI [0.45; 4.71], $P(\delta > 0) = 0.98$); no robust difference to NN was found. On Syll4, synchrony is higher in NG than both GN ($\delta = 6.23$, CI [4.40; 8.06], $P(\delta > 0) = 1$) and NN ($\delta = 7.03$, CI [5.21; 8.84], $P(\delta > 0) = 1$).

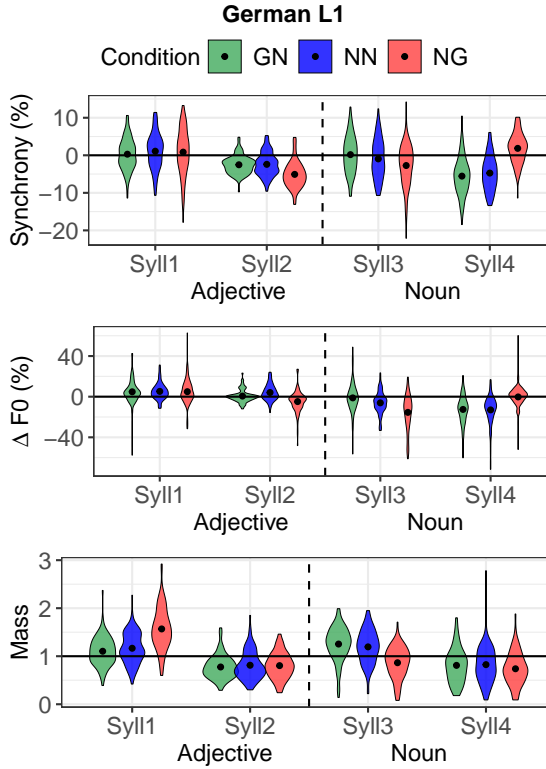


Figure 2.14: Aggregated values of synchrony, $\Delta F0$ and mass (on the y-axes) pooled across German native speakers. The x-axis displays the four syllables of the noun phrases, with Syll1 and Syll2 being the adjective and Syll3 and Syll4 the noun. Information structure conditions are colour-coded: green for given-new (GN), blue for new-new (NN) and red for new-given (NG).

- $\Delta F0$ on Syll3 is lower in NG than in GN ($\delta = 13.55$, CI [6.97; 20.09], $P(\delta > 0) = 1$) and NN ($\delta = 8.74$, CI [2.05; 15.14], $P(\delta > 0) = 0.99$). On Syll4, $\Delta F0$ is higher in NG than GN ($\delta = 11.83$, CI [7.05; 16.47], $P(\delta > 0) = 1$) and NN ($\delta = 12.31$, CI [7.26; 17.59], $P(\delta > 0) = 1$).
- *Mass* on Syll3 is lower in NG than GN ($\delta = 0.37$, CI [0.25; 0.50], $P(\delta > 0) = 0.99$) and NN ($\delta = 0.31$, CI [0.19; 0.44], $P(\delta > 0) = 0.99$). Within GN Syll3 is higher than Syll1 ($\delta = 0.13$, CI [0.01; 0.27], $P(\delta > 0) = 0.97$), while the opposite is true for NG, with Syll1 being higher than Syll3 ($\delta = 0.69$, CI [0.53; 0.86], $P(\delta > 0) = 1$). In NN, mass is highly similar across Syll1 and Syll3 ($\delta = 0.01$, CI [-0.11; 0.14], $P(\delta > 0) = 0.60$).

The models confirm the patterns identified through data visualisation and, in line with previous findings, show that prosodic marking of information status in native German is tendentially achieved via high F0 and strong energy. On average, GN and NN present a similar intonation contour, whereas the contour for NG differs, reflected in distinct values of synchrony and $\Delta F0$ on Syll3 and Syll4. Moreover, stronger energy corresponding to new information status is mirrored in mass values across all conditions, with the stressed syllable of the new element being strong and relatively stronger than the given element within the same condition.

2.5.1 Summary

The results are in line with previous studies on L1 German, showing that prosodic marking of information status is achieved by different modulations of F0 and energy across new and given elements. Despite some variability, overall F0 starts falling after a new element, in the form of either an F0 peak on the new element (typically when the first element is new, i.e. in NG), or a high plateau until the new element (when the second element is new, i.e. in GN and NN). Energy patterns reflect information status contrasts across all conditions (i.e. stronger energy corresponding to the stressed syllable of new elements as compared to given elements), and stress contrasts only in GN and NN conditions (i.e. stressed syllables displaying strong energy and unstressed syllables weak energy). Stress-related distinctions do not hold for NG, showing that, in the presence of post-focal given material, information status marking overrides stress marking and prosodic attenuation takes place, in line with the phenomenon of deaccentuation widely described in the literature.

2.6 Results: German L2

Figure 2.15 displays three example periograms for the three information structure conditions uttered by Italian learners of German in their L2. The periograms show two different F0 patterns across conditions. In the GN and NN conditions, F0 is rising throughout the first syllable, reaches a peak on the second syllable and finally falls on the third syllable. In NG the peak is fully realised on the first syllable, where the falling movement takes place as well. As a consequence of the different location of the F0 peak within the first word, the contour shapes are dissimilar across the two words, that is between the second and the third syllable. This position displays a fall in the GN and NN conditions, while in NG it is quite flat, with the fall already completed beforehand. Therefore, the presence of new

vs. given information in the final position of the noun phrase is characterised by two different F0 patterns distinguished by the position of the F0 peak: later in the first word in GN and NN, and earlier in the first word in NG.

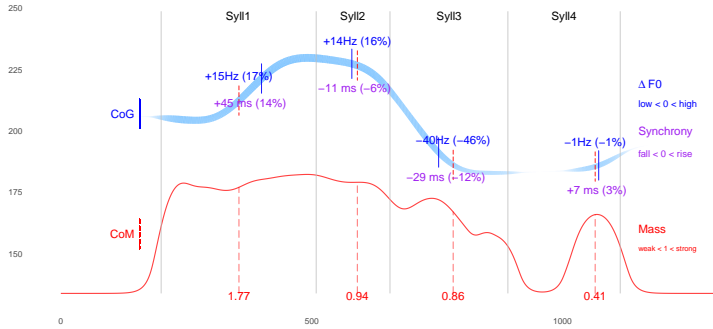
The aggregated data of synchrony, $\Delta F0$ and mass in Figure 2.16 confirm these observations. In greater detail:

- *Synchrony*. Positive distributions on Syll1 of GN and NN reflect a rising F0 trajectory. In contrast, predominantly negative values on Syll1 reflect the earlier F0 fall in NG. The effect of the later F0 fall in GN and NN is reflected in more negative synchrony values on Syll3 than in NG.
- $\Delta F0$. Positive values on Syll2 and negative values on Syll3 are indicative of the F0 peak on Syll2 in GN and NN. In contrast, negative values already on Syll2 in NG indicate that the F0 peak is located within Syll1, resulting in less negative values on Syll3 as compared to GN and NN.
- *Mass*. All conditions present the same pattern and display strong energy on Syll1 and weak energy on Syll3. This pattern does not reflect stress-related distinctions as Syll3 is stressed, nor information status contrasts between NG and GN. Still, the two conditions do not seem to be totally equal and there is a subtle difference in the shape of the distributions on the stressed syllables (Syll1 and Syll3) across NG and GN. It seems that Syll1 in NG (stressed syllable of the new word) is stronger than in GN (stressed syllable of the given word), whereas Syll3 in NG is weaker (stressed syllable of the given word) than in GN (stressed syllable of the new word).

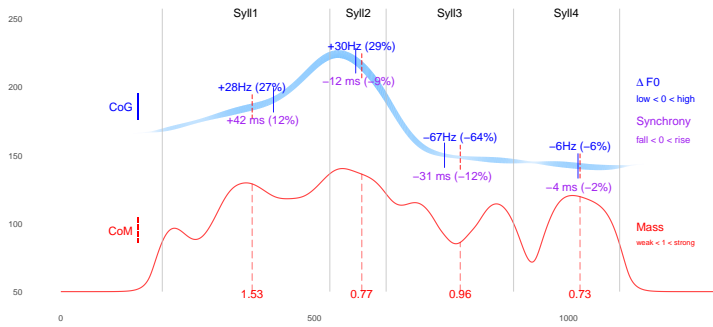
Results of the Bayesian analysis support these observations, including those of the more subtle modulations of mass values:

- *Synchrony* on Syll1 is higher in GN ($\delta = 4.44$, CI [3.39; 5.62], $P(\delta > 0) = 1$) and NN ($\delta = 5.41$, CI [4.22; 6.48], $P(\delta > 0) = 1$) than in NG. In contrast, synchrony on Syll3 is lower in GN ($\delta = 1.62$, CI [0.52; 2.78], $P(\delta > 0) = 0.99$) and NN ($\delta = 2.20$, CI [1.06; 3.36], $P(\delta > 0) = 1$) compared to NG.
- $\Delta F0$ on Syll2 is higher in GN ($\delta = 19.52$, CI [13.83; 25.73], $P(\delta > 0) = 1$) and NN ($\delta = 20.12$, CI [14.3; 26.66], $P(\delta > 0) = 1$) than in NG. In contrast, $\Delta F0$ on Syll3 is lower in GN ($\delta = 7.15$, CI [2.99; 11.66], $P(\delta > 0) = 0.99$) and NN ($\delta = 8.20$, CI [3.8; 12.47], $P(\delta > 0) = 1$) than in NG.
- *Mass* on Syll1 is lower for GN ($\delta = 0.10$, CI [0.05; 0.15], $P(\delta > 0) = 1$) and NN ($\delta = 0.07$, CI [0.02; 0.11], $P(\delta > 0) = 0.99$) compared to NG. On Syll3 GN ($\delta = 0.05$, CI [0.006; 0.10], $P(\delta > 0) = 0.98$) is higher than NG.

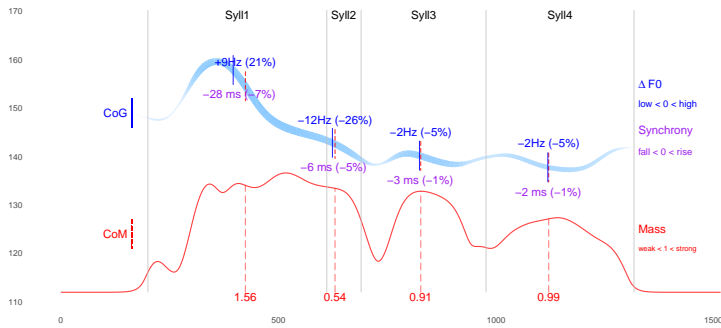
2 Prosodic marking of information status



(a) GN. The token is *graue Vase* (Eng. 'grey vase').



(b) NN. The token is *braune Dose* (Eng. 'brown can').



(c) NG. The token is *braune Birne* (Eng. 'brown pear').

Figure 2.15: Example perioqrams displaying F0 and periodic energy for two-word noun phrases in three information structure conditions produced by Italian learners of German as L2.

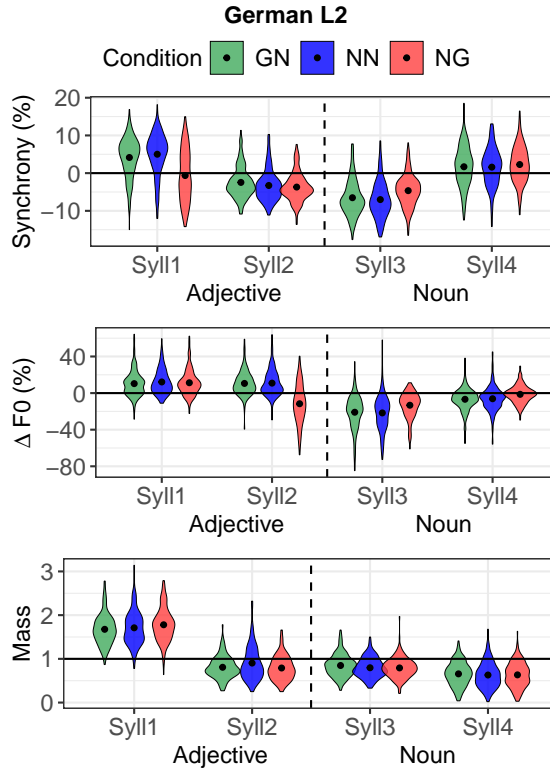


Figure 2.16: Aggregated values of synchrony, $\Delta F0$ and mass (on the y-axes) pooled across German L2 learners. The x-axis displays the four syllables of the noun phrases, with Syll1 and Syll2 being the adjective and Syll3 and Syll4 the noun. Information structure conditions are colour-coded: green for given-new (GN), blue for new-new (NN) and red for new-given (NG).

Statistical results support the data showing two different F0 contours distinguished by the position of the F0 peak: within the first syllable in NG and later in the first word in GN and NN. One pattern of modulation for mass across all conditions was found, namely strong mass on the first word and weak mass on the second word, which pattern matches information status contrasts only in the NG condition.

2.6.1 Proficiency levels

Since learners had different proficiency levels, it is interesting to assess whether and to which extent their ability to mark information status prosodically im-

proves in the learning process, that is, across proficiency levels. To this end, learners were categorised into two main groups, i.e. beginner and advanced learners. In the previous section, it emerged that GN and NN present similar patterns for the three measures of synchrony, $\Delta F0$ and mass, in all language groups. Therefore, when exploring results across proficiency levels and, successively, across language groups, I will only comment on the difference between GN and NG and leave out the all-new condition to simplify the presentation of the results and increase clarity.

Figure 2.17 shows that both proficiency groups produce the same patterns for F0 and mass modulation. Subtle differences across proficiency levels can be spotted, with beginners displaying less discrimination between conditions as compared to advanced learners. In particular:

- *Synchrony*. In the beginner group, the NG distribution on Syll1 presents a similar amount of positive and negative values, whereas the distribution tends to be more negative in the advanced group, showing a better discrimination of the earlier F0 fall in the NG condition. On Syll3, while mean values for NG and GN appear to be similar across proficiency groups, the distribution of data is less spread out in the advanced compared to the beginner group.
- $\Delta F0$. In the beginner group, values on Syll2 and Syll3 tend to be closer to zero for both conditions than in the advanced group, in which values for GN and NG tend towards opposite directions. As a result, beginners present lower values for GN and higher values for NG than advanced learners on Syll2, while on Syll3, beginners present higher values for GN than advanced learners. This shows that beginners use a reduced range when modulating F0 across syllables and conditions.
- *Mass*. The subtle contrasts regarding information status on Syll1 and Syll3 seem to be enhanced in the advanced group. In particular, for GN Syll1 is lower and Syll3 higher than in the beginner group.

These observations were confirmed to be statistically robust trends. A Bayesian analysis was used to test all relevant contrasts for the measures of synchrony, $\Delta F0$ and mass within groups (i.e. beginners and advanced, see bullet points in Section 2.6), and across groups (see correspondent bullet points above), in order to assess: 1) whether the contrasts among conditions hold true within both proficiency groups and 2) whether the two groups robustly differ from each other.

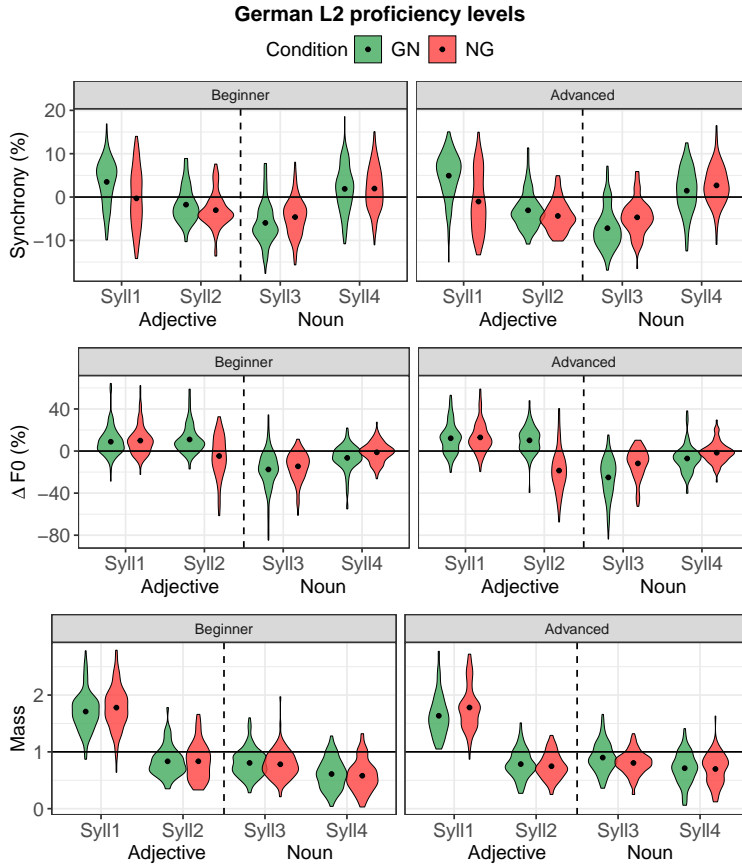


Figure 2.17: Aggregated values of synchrony, $\Delta F0$ and mass (on the y-axes) pooled across German L2 learners by proficiency levels. Beginners are on the left and advanced learners on the right. The x-axis displays the four syllables of the noun phrases, with Syll1 and Syll2 being the adjective and Syll3 and Syll4 the noun. Information structure conditions are colour-coded: green for given-new (GN) and red for new-given (NG).

- *Synchrony*. NG and GN are robustly different on Syll1 within both proficiency groups, confirming the general tendencies (for beginners: $\delta = 3.64$, CI [2.21; 5.07], $P(\delta > 0) = 1$; for advanced: $\delta = 5.6$, CI [3.95; 7.14], $P(\delta > 0) = 1$) and on Syll3 only within the advanced group ($\delta = 2.3$, CI [0.70; 3.95], $P(\delta > 0) = 0.99$). Across groups, there is compelling evidence only for a difference in GN on Syll1, which is higher in the advanced group ($\delta = 1.38$, CI [-0.10; 3.05], $P(\delta > 0) = 0.95$). However, Syll3 in GN shows a relatively high difference across groups ($\delta = 1.30$, CI [-0.68; 3.27], $P(\delta > 0) = 0.90$) and is lower in the advanced group.
- $\Delta F0$. NG and GN are robustly different on Syll2 within both proficiency groups, confirming the general tendencies (for beginners: $\delta = 13.54$, CI [6.02; 21.67], $P(\delta > 0) = 0.99$; for advanced: $\delta = 26.95$, CI [19.36; 34.74], $P(\delta > 0) = 1$) and on Syll3 only within the advanced group ($\delta = 12.7$, CI [6.08; 18.67], $P(\delta > 0) = 0.99$). Across groups, there is strong evidence for a difference in NG on Syll2 ($\delta = 13.03$, CI [3.11; 22.87], $P(\delta > 0) = 0.99$) and in GN on Syll3 ($\delta = 7.37$, CI [0.52; 14.2], $P(\delta > 0) = 0.98$), which are lower in the advanced group.
- *Mass*. NG and GN are robustly different on Syll1 within both proficiency groups, confirming the general tendencies (for beginners: $\delta = 0.07$, CI [0.01; 0.13], $P(\delta > 0) = 0.98$; for advanced: $\delta = 0.14$, CI [0.07; 0.21], $P(\delta > 0) = 0.99$) and on Syll3 only within the advanced group ($\delta = 0.08$, CI [0.01; 0.16], $P(\delta > 0) = 0.99$). Across proficiency levels, there is a relatively high difference in GN on Syll3 ($\delta = 0.07$, CI [-0.01; 0.17], $P(\delta > 0) = 0.94$), which is higher in the advanced group.

The models suggest that most of the between-group differences concern the modulation of synchrony and $\Delta F0$ values on the first three syllables. These slight variations across groups seem to suggest that with increasing proficiency, learners better discern the two pragmatic conditions, enhancing a distinct modulation of their phonetic details, although this does not result in different patterns in beginner as compared to advanced learners.

2.6.2 Summary

In both their L2 and their native language, learners distinguish two intonation patterns through the modulation of the $F0$ falling movement on the first word, showing a transfer of L1 Italian patterns. As in their native language, they use an $F0$ peak aligned near the onset of the first syllable to mark the NG condition

2.7 Interlanguage compared to learners' native and target languages

(i.e. earlier on the first word), whilst they realise the F0 peak across the first two syllables in the GN and NN conditions (i.e. later on the first word). However, learners use energy in a way that is different from both L1 Italian and L1 German. They produce the same pattern in all pragmatic conditions: the first syllable, which is lexically stressed, is strong while the rest of the noun phrase displays a reduction in energy. This shows that energy is never used according to lexical stress patterns and only in the case of NG signals information status differences within the noun phrase.

The analysis by proficiency level has shown that robust differences between beginner and advanced learners are present, but that they are limited and not consistent across measures, syllables, and conditions. Specifically, the advanced group seems to enhance the difference between their NG and GN realisations more as compared to beginners. In contrast, beginners consistently differentiate the two conditions less clearly on the second word by neutralising F0 and energy distinctions across conditions. Most importantly, these differences do not result in overall distinct patterns, and the two proficiency levels still show the same trends, which is an expected result considered that prosody is not explicitly thematised in L2 classrooms. For this reason, I will consider them as one single group in the next section, i.e. when comparing L2 German to L1 Italian and L1 German.

2.7 Interlanguage compared to learners' native and target languages

After having shown in detail the different strategies used by the three groups to prosodically mark information status, it is time to compare learners' prosodic realisations of new-given (NG) and given-new (GN) information structures with their native baseline, Italian, and target language, German⁵ To do so, a preliminary discussion of the main cross-linguistic differences and similarities between the two native languages is necessary to highlight the relevant aspects for learners (see Rasier & Hiligsmann 2007 for a review of models of second language acquisition). In light of this cross-linguistic comparison, I will then discuss learners' productions.

⁵In the previous chapters, it emerged that GN and NN present similar F0 and energy patterns in all language groups. Thus, to break down complexity, I will only comment on the difference between GN and NG and leave out the NN condition. All data are available and can be inspected in the OSF repository (<https://osf.io/9ca6m/>).

2.7.1 F0 contours

For a direct comparison, Figure 2.18 displays the averaged F0 contours found for the realisation of NG and GN in the three different language groups, i.e. Italian L1, German L2 and German L1.

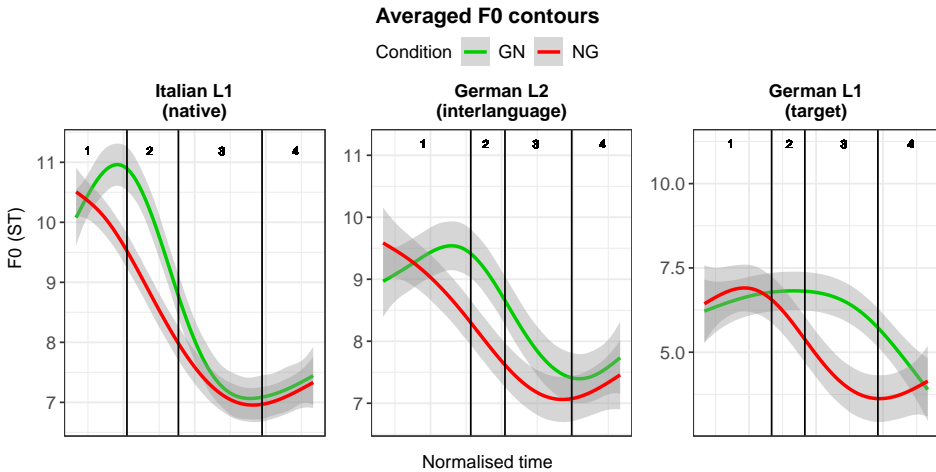


Figure 2.18: Averaged F0 contours pooled across speakers for each language group. The y-axis shows F0 in semitones, while the x-axis shows normalised time aligned at the boundary between the two words of the noun phrase. Syllables of the noun phrase are numbered from one to four and syllable boundaries are marked by vertical black lines. The grey area around the contours represents the standard error and contours are colour-coded according to their information structure condition: green for given-new (GN) and red for new-given (NG).

Comparing native Italian with native German, it is apparent that the two languages use different strategies to express the contrast between the two pragmatic conditions: L1 Italian prosodically differentiates the two conditions on the first word, while L1 German does so on the second word of the noun phrase. Specifically, in L1 Italian, NG shows an F0 fall early in the first syllable and GN on the second syllable so that in both cases the peak is located within the first word. In L1 German, in contrast, NG has an F0 fall on the second syllable with a peak on the first syllable and GN has a hat pattern, with the F0 fall on the third syllable and high F0 stretching across the first and the second word.⁶ Thus, from an Ital-

⁶Even if averaged data for L1 German displays a hat pattern for GN, this does not mean that this contour is the one most used for this condition. Contours with a peak either on the first or on the second element contribute to the mean (see examples in Section 2.5). This can be seen by considering aggregated data for synchrony on Syll2, showing a tendency for slightly falling F0 (Figure 2.14) as well as by looking at by-speaker contours (Figure A3 in the Appendix).

ian learner's perspective, the timing of the F0 fall in German is later than in their native baseline.

Learners seem to transfer their L1 F0 shapes to their L2 realisations, as the intonation patterns in the L1 and L2 appear to be quite similar,⁷ even though they exploit a reduced F0 range as compared to their baseline (similar to the target language, which will be further discussed in Section 2.7.3). This means that learners still distinguish information status within noun phrases, but they do so using a different timing of the F0 fall on the first word and, therefore, do not match the target contours produced by German native speakers. In particular, in NG, learners' F0 fall takes place too early in the first word compared to the target (on the first syllable instead of the second syllable) and, in GN, learners do not produce a hat pattern across the two words and the F0 fall occurs before the new element (on the second syllable instead of the third syllable). As a result, differences between learners' realisations and their target language are evident on the first word in the NG condition and on the second word in the GN condition.

Figure 2.19 provides greater detail on F0 modulation across the three language groups, showing again values of synchrony and $\Delta F0$ for the relevant locations in the noun phrases in light of the differences observed, i.e. Syll1 and Syll3 for synchrony and Syll2 and Syll3 for $\Delta F0$. Distributions and mean values of synchrony and $\Delta F0$ for the learner group are midway between their native and target languages (with the exception of synchrony on Syll1 in NG, displaying more negative values than either L1). A Bayesian analysis confirms the robustness of this observation:

- *Synchrony*. On Syll1 in NG, values for the L2 group are lower than both in the native ($\delta = 1.09$, CI [0.31; 1.81], $P(\delta > 0) = 0.99$) and the target language ($\delta = 1.30$, CI: [0.29; 2.19], $P(\delta > 0) = 0.99$). On Syll3 in GN, values for the L2 group are higher than in the native ($\delta = 2.66$, CI [1.83; 3.43], $P(\delta > 0) = 1$) and lower than in the target language ($\delta = 6.02$, CI: [5.02; 7.04], $P(\delta > 0) = 1$).
- *$\Delta F0$* . On Syll2 in NG, values for the L2 group are higher than in the native language ($\delta = 5.01$, CI [2.3; 7.77], $P(\delta > 0) = 1$), but still lower than in the target language ($\delta = 3.49$, CI [0.03; 7.01], $P(\delta > 0) = 0.97$). The same holds true for Syll3 in GN (difference to L1 Italian: $\delta = 13.89$, CI [11.57; 16.24], $P(\delta > 0) = 1$; difference to L1 German: $\delta = 16.56$, CI [13.46; 19.81], $P(\delta > 0) = 1$).

⁷Notice that the difference in duration of the first syllable across languages depends on the underlying segmental material, i.e. a different syllabic structure. At the beginning of German NPs there are words like *graue* and *blaue* with two consonants in the onset and a diphthong as the nucleus, while in Italian all syllables are composed of a single consonant and a vowel.

2 Prosodic marking of information status

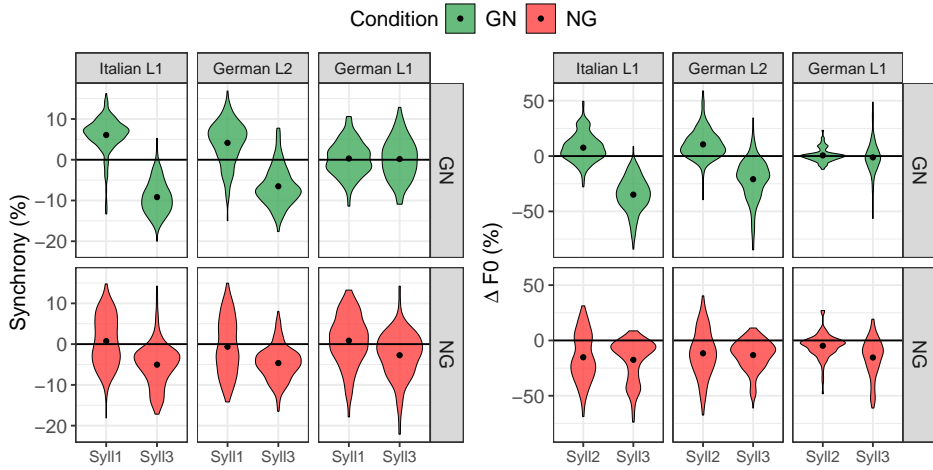


Figure 2.19: Aggregated values of synchrony and $\Delta F0$ for the relevant syllables across language groups. Synchrony values are displayed for syllables one and three (left panels), while $\Delta F0$ values are shown for syllables two and three (right panels). Information structure conditions are colour-coded and positioned on two separate rows: green for given-new (GN, upper row) and red for new-given (NG, bottom row).

Despite the fact that learners seem to transfer their L1 contours to their L2, the models provide strong evidence for some differences in their continuous modulation. In particular, in learners' interlanguage, there is a less steep slope on the first word in the NG condition and a narrower F0 range overall as compared to their native language, with the latter being a feature of native German as well. Still, learners' F0 patterns mirror their Italian native productions and do not resemble the target ones.

2.7.2 Mass

The previous measures give an overview of continuous parameters that quantify what is visible in the averaged contours. However, this provides little information about possible accentuation. Thus, I will now compare the use of periodic energy mass across language groups, as mass is derived from a calculation accounting for power and duration, two parameters which are often involved in accentuation together with F0 movement as described in Section 2.1.2. I will focus on the third syllable, as that is the one where deaccentuation should be realised by learners according to pragmatic condition in order to match the accentuation patterns of the target language.

2.7 Interlanguage compared to learners' native and target languages

Values of mass for the third syllable of the noun phrase are shown in Figure 2.20. Focussing first on native Italian and native German, the results provide evidence in line with the literature (Section 2.1.4 for Italian, Section 2.1.5 for German). L1 Italian displays strong mass on Syll3, both when the second word of the noun phrase is new and when it is given and post-focal, supporting the finding that the final word requires an accent, independent of pragmatic status. By contrast, in L1 German there is strong mass only on Syll3 when the second word of the noun phrase is new, whereas mass is weak on the post-focal given element, showing that, in line with previous studies reporting deaccentuation, the nuclear position is prosodically highlighted or attenuated according to information status.

In learners' L2, mass is clearly weak across conditions, with values similarly distributed below one. The values appear to be even more negatively distributed than in the NG condition by L1 German speakers, also displaying weak mass. As a result, learners do not seem to transfer mass patterns from their native language as they do for F0 and, instead, show prosodic attenuation as in the target language, which might be interpreted as an attempt to reproduce deaccentuation. However, in comparison with L1 German, learners' mass on Syll3 is weak not only in NG, but also in GN, i.e. they appear to deaccent new information as well.

The robustness of the differences observed across language groups were confirmed by Bayesian models:

- *Mass.* On Syll3 in both conditions, the L2 group presents lower values than both in the native (difference for NG: $\delta = 0.34$, CI [0.29; 0.39], $P(\delta > 0) = 1$; and GN: $\delta = 0.38$, CI [0.33; 0.43], $P(\delta > 0) = 1$) and target language (difference for NG: $\delta = 0.09$, CI [0.03; 0.15], $P(\delta > 0) = 0.99$; and GN: $\delta = 0.42$, CI [0.35; 0.48], $P(\delta > 0) = 1$).

Model results confirm that learners' prosodic strength patterns diverge from their native Italian and tend to reproduce the native German for the NG condition, with the difference that 1) they attenuate the post-focal element even more than native speakers of German and 2) this same pattern is extended to GN as well, showing that they do not use prosodic strength to mark information status contrasts as in their native Italian.

2.7.3 Discussion

The production of two-word noun phrases with different information status, new-given (NG) vs. given-new (GN), in L2 German spoken by Italian learners was compared to their native language and the target language.

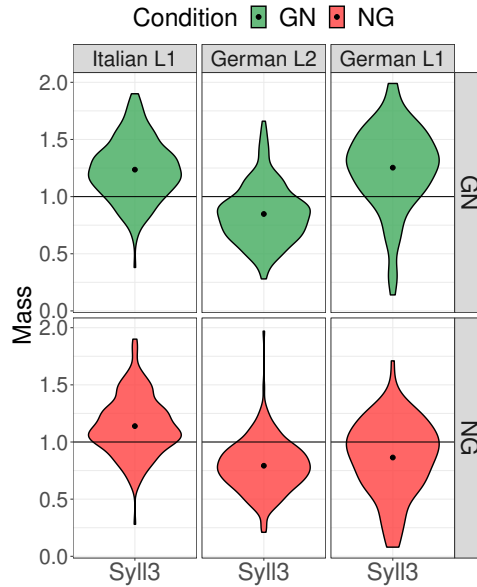


Figure 2.20: Aggregated values of mass for syllable three across language groups. Information structure conditions are colour-coded and positioned on two separate rows: green for given-new (GN, upper row with Syll3 being a new item) and red for new-given (NG, bottom row with Syll3 being a given item).

Results show that Italian learners of German differentiate the two pragmatic conditions using two F0 shapes which highly resemble their native Italian ones, i.e. by producing the F0 peak within the first syllable in NG and later in the first word in GN. However, statistical results provide evidence against a complete transfer, showing a systematically different modulation of F0. Specifically, learners produce the anticipated peak in NG with a less steep slope and use a reduced F0 range across conditions. A narrower F0 range is characteristic of L1 German, but the hypothesis that learners intentionally compress their Italian native F0 range to approach the target language is contestable, as other L2 studies involving language pairs different from the one object of this study have made the same observation (Dutch learners of Greek in Mennen 1998; Taiwan Mandarin learners of English in Visceglia et al. 2011; Italian learners of English in Urbani 2012; French learners of German and German learners of French in Zimmerer et al. 2014; Chinese learners of Japanese in Shi et al. 2014). This has been interpreted as a characteristic of interlanguages, and not necessarily as ascribable to transfer or learning effects, but possibly to insecurity in speaking a second language (Mennen 1998; Shi et al. 2014; Zimmerer et al. 2014).

It was also found that, as in their L1, Italian learners of German do not make use of prosodic strength to distinguish the two pragmatic conditions. However, the pattern learners use is not present in their L1 and, instead, might be interpreted as a possible attempt to reproduce a pattern typical of German. In particular, learners strengthen the first word and weaken the second, similarly to L1 German productions in the NG condition. Hereby, learners enhance the attenuation of the second word even more than L1 German speakers, probably as a form of hypercorrection, which has already been documented in L2 phonetic acquisition (cf. Eckman et al. 2013; Kelly 2022; Petrov 2021). Learners apply this energy pattern across all pragmatic conditions, possibly because in their L1 Italian, prosodic strength is not used to encode information structure (but only positional rules, accenting the final position independently of IS), and they may not be aware of the different function it serves in German. For these reasons, instead of using prosodic attenuation to convey information status, they seem to perceive and reproduce it as a salient feature of native German speech according to positional rules (deaccenting the nuclear position independently of IS).

2.7.4 Summary

Italian learners of German produce F0 shapes that resemble those in their L1. However, they use a different energy pattern across all conditions, one that is similar to L1 German in the NG condition and can be interpreted as an attempt to reproduce target-like productions. Nevertheless, modelling the phonetic details reveals that learners' modulations of F0 and energy are significantly different from both their native Italian baseline and the native German target. This indicates that although the patterns appear similar, they are phonetically implemented in different ways, i.e. learners do not fully transfer features from their L1 nor achieve complete alignment with their target language. Instead, they exhibit the expected characteristics of an interlanguage.

2.8 Implications for phonological analysis

The majority of recent studies on the mapping between prosody and information status have been carried out within the Autosegmental-Metrical (AM) phonology framework. For this reason, a critical discussion of the present findings based on a phonetic periodic-energy-based approach in relation to the existing literature rooted in a phonological framework is necessary. To confront the different theoretical frameworks and validate the use of this new periodic-energy-based analysis in L2 research, I will compare periodic-energy-related measures to the

established measures of alignment, scaling, duration, and intensity, and suggest a possible description of the contours found in terms of categorical pitch accent types. The goal is twofold: to critically discuss the advantages and disadvantages of a phonetic and phonological approach in L2 research and to validate the use of a periodic-energy-based method, especially for L2 analyses.

2.8.1 Background

Most recent studies on the prosodic marking of information status have been conducted within the framework of AM phonology, using the ToBI (Tones and Break Indices) system for intonation transcription. The system was initially based on American English (Beckman & Pierrehumbert 1986; Beckman & Ayers 1997; Beckman et al. 2005; Veilleux et al. 2006) and has been fully adapted for German (GToBI: Grice et al. 2005) but only sketched for some Italian varieties (Grice et al. 2005).

This model sees F0 turning points as tonal targets joined by quasi-linear interpolation. These targets are phonologically represented as tones, with abstract discrete values: H for a high target, also called *peak*, and L for a low target, also called *valley*. In German ToBI (Grice et al. 2005), there are two further operators for H tones to describe the relative height of the target: *downstep*, that is a lower H target compared to the previous H target and is transcribed with an exclamation mark (!H), and *upstep*, that is a higher H target compared to the previous H target, and is transcribed with a caret (^H). The *metrical* aspect of the model is reflected in the division of utterances into phrases and the assignment of relative prominence to the elements within the phrases, while the *autosegmental* aspect refers to the fact that tonal and segmental features are considered to be independent and are indeed represented on different, parallel and autonomous levels. The elements on the different levels are then connected by associations at specific points, leading to the anchoring of the tune to the text on single syllables and/or at the edges of phrases. The resulting types of phonological categories are defined as pitch accents and edge or boundary tones, respectively. Pitch accents mark prominence and are typically associated with metrically strong syllables and transcribed with a star (e.g. L*; H*). Edge or boundary tones mark the edges of constituents and are associated with their periphery (Grice et al. 2005). They are transcribed with a “-” for edges of minor phrases, also called *intermediate phrases*, and with a “%” for edges of major phrases, also called *intonation phrases*.

In the development of AM phonological systems, gradual phonetic parameters have been investigated for their contribution towards establishing phonological categories. Pitch accent categories are often determined on the basis of

segmental anchoring of F0 turning points, which involves temporal alignment or synchronisation of F0 target peaks and valleys to landmarks in the segmental string (horizontal axis), such as the CV boundary of an accented syllable (i.e. the onset of the vowel, see Arvaniti et al. 1998). Alongside alignment, other gradual parameters typically used to describe pitch accent categories are scaling (vertical axis) – the relative F0 height of the tonal target encoded in accent type, H^* being typically lower than $L+H^*$ – and pitch excursion – the direction (rising vs. falling) and extent (small vs. large) of the excursion resulting from the scaling and height of pitch and the alignment of a pitch peak or valley with a stressed syllable. Moreover, for German the *tonal onglide* is of particular importance (Ritter & Grice 2015). This describes the F0 movement from the preceding syllable to the tonal target on the accented syllable, often encoded in the leading tone of pitch accents: $L+H^*$ for a rising onglide from an L to an H target or $H+L^*$ for a falling onglide from an H to an L target.

In the following sections I present an alternative analysis of the collected data using the AM approach. This includes the use of alignment and scaling to explore F0 contours, duration and intensity to examine accentuation patterns, and providing a phonological interpretation using ToBI. The results are discussed in light of my findings based on a periodic-energy-based approach.

2.8.2 Alignment and scaling

In the study I conducted, the measures of synchrony and $\Delta F0$ were used to analyse differences in F0 modulations used to mark information status contrasts. Within the AM framework, alignment and scaling are the two most widely used measures for investigating the phonetic details of tonal targets and describe the F0 contour. Alignment reflects the temporal coordination of tonal targets within the segmental string and scaling reflects their pitch level (for an extensive overview see Chapter 5 in Ladd 2008).

To provide a comparative analysis of F0 shapes using these two measures (Cangemi et al. 2019), alignment is extracted as the time point corresponding to the highest F0 point of rising-falling F0 contours and normalised by token duration (thus, it is represented on a time scale from 0 to 1). Scaling is extracted in Hz as the highest F0 point of rising-falling F0 contours and normalised by each speaker's range. Representing the alignment of F0 peaks realised within the target noun phrases implies that only rising-falling F0 shapes realised on and around the region of a syllable are accounted for, i.e. F0 contours that are simply rising or simply falling are not included in the analysis.

Figures 2.21–2.23 display values of normalised alignment and scaling of F0 peaks realised within the noun phrases under the three information structure conditions: GN (given-new), NN (new-new) and NG (new-given).

Figure 2.21 shows alignment and scaling values for L1 Italian data, with all items visible in the upper panel and only items correctly matched to information status in the lower panel. As already observed in the periodic-energy based analysis (Section 2.3), NG forms a well-separated cluster of data points compared to the other two conditions, which, instead, overlap. This trend is clearer in the lower panel, after filtering out the items which were not identified as corresponding to the intended information structure in the perceptual rating task (Section 2.4). Thus, both methods show that two intonation contours with a different peak alignment within the first word distinguish the information status of the following word, that is, post-focal given corresponding to early alignment on the first word vs. new corresponding to late alignment on the first word.

Figure 2.22 shows values for alignment and scaling for L1 German, from which conclusions similar to those in the periodic-energy-based analysis can be derived. In NG, the F0 peak is consistently aligned between the first and the second syllable, that is, on the new word. GN and NN show instead more variability (Section 2.5), with a wide distribution of values showing that the peak can be located either late on the first word or on the second word. However, it is necessary to point out that this figure only shows a reduced subset of data relative to the specific F0 configuration of peaks (42% of the occurrences). Other configurations found, such as the hat pattern or the double peak, cannot be represented using this measure without making an arbitrary decision as to where the peak is labelled (following Welby 2004).

The transfer of native F0 contours by Italian learners to their L2 German, assessed by a periodic-energy-based analysis (Section 2.6), is also visible using alignment and scaling. Figure 2.23 shows that, similar to their Italian L1, distributions of data points in L2 German tend to form two clusters, separating the F0 peak of NG, which is aligned earlier in the first word, from the F0 peak of GN and NN, which is aligned later in the first word. Moreover, learners do realise the F0 peak on the second word for some GN and NN instances as in native German, but this is rarely the case as shown by the few green (GN) and blue (NN) data points on Syll3 and Syll4 (second word) as compared to the much more numerous ones on Syll1 and Syll2 (first word).

To summarise, even using the established measures of alignment and scaling to quantify the F0 shape, the same F0 patterns identified via synchrony and $\Delta F0$ for all language groups clearly emerge. This confirms and validates our results based on periodic-energy-based measures and can be summarised as follows:

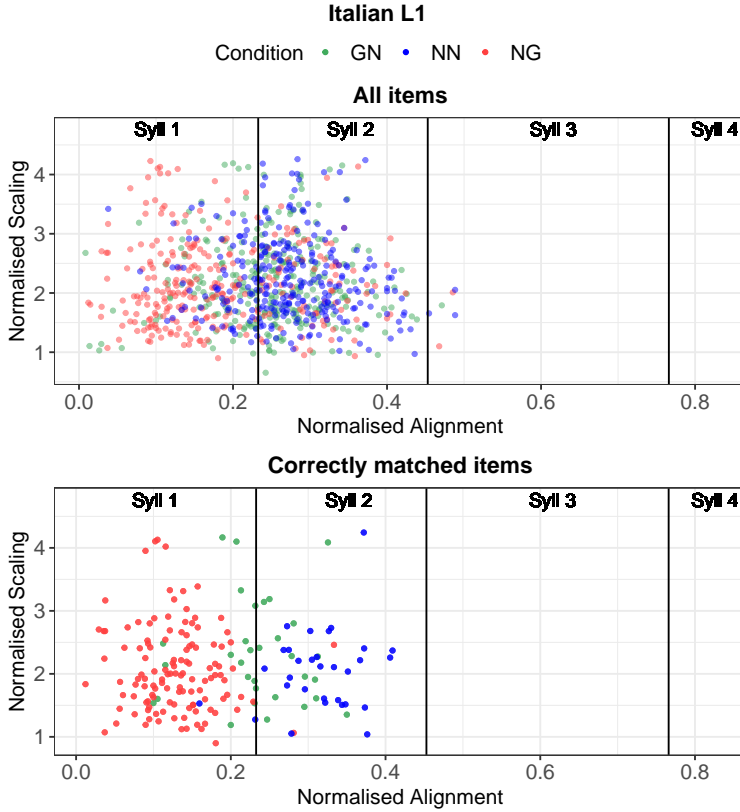


Figure 2.21: Alignment and Scaling for L1 Italian data (all data, on top, vs. correctly matched items only, on the bottom). Mean time-normalised boundaries of the four syllables of the noun phrases are marked by solid vertical lines. Syll1 and Syll2 form the noun and Syll3 and Syll4 form the adjective. Information structure conditions are colour-coded: green for given-new (GN), blue for new-new (NN) and red for new-given (NG).

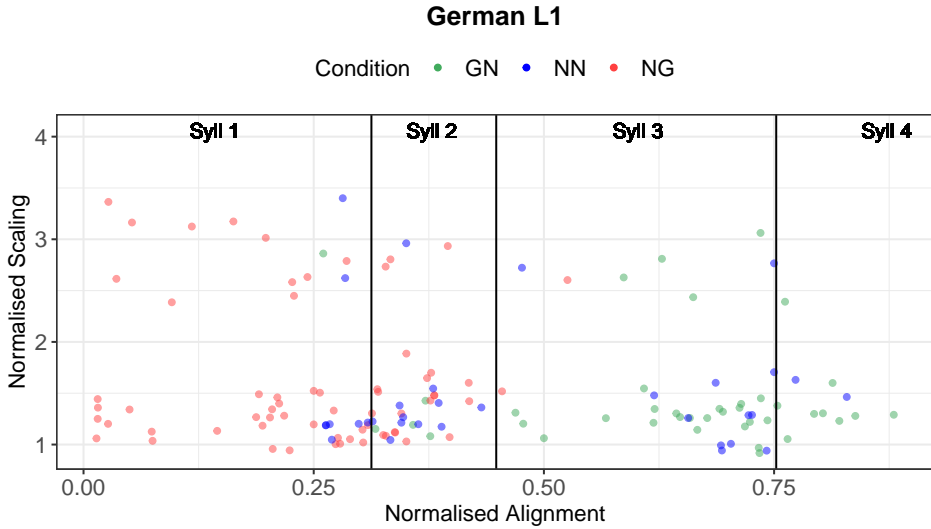


Figure 2.22: Alignment and Scaling for L1 German data. Mean time-normalised boundaries of the four syllables of the noun phrases are marked by solid vertical lines. Syll1 and Syll2 form the adjective and Syll3 and Syll4 form the noun. Information structure conditions are colour-coded: green for given-new (GN), blue for new-new (NN) and red for new-given (NG).

- In L1 Italian, the position of the F0 peak on the first word distinguishes the information status of the second word;
- In L1 German, the F0 peak typically occurs on the new element. This is very consistent in NG and more variable in GN and NN information structures, for which other contour shapes are also realised;
- In L2 German, Italian learners reproduce their L1 patterns (with very few exceptions of target-like F0 patterns).

2.8.3 Duration and Intensity

Periodic energy mass, derived from a calculation accounting for duration and power, was used in the present study to analyse prosodic strength patterns as a cue to accentuation, while two well-established measures used to investigate power and length are intensity and duration. For a comparison with mass, the distributions of values of mean intensity (in decibels) and duration (in seconds) are presented for the lexically stressed syllables (Syll1 for the first word, Syll3 for the second word) in the GN and NG conditions for each language group.

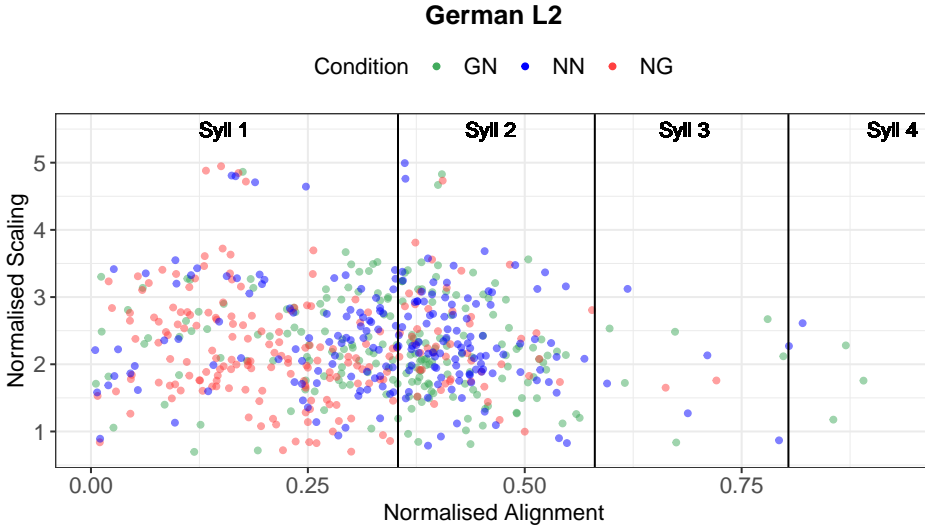


Figure 2.23: Alignment and Scaling for L2 German data. Mean time-normalised boundaries of the four syllables of the noun phrases are marked by solid vertical lines. Syll1 and Syll2 form the adjective and Syll3 and Syll4 form the noun. Information structure conditions are colour-coded: green for given-new (GN), blue for new-new (NN) and red for new-given (NG).

Differences among conditions and syllables were statistically tested using Bayesian hierarchical linear models,⁸ with the output reported in parentheses throughout the description of the results.

In L1 Italian (Figure 2.24), distributions of intensity and duration across the noun phrase seem to each enhance a different word, yielding conflicting results. For both conditions, duration shows that Syll3 is longer than Syll1 (for GN: $\delta = 0.07$, CI [0.07; 0.08], $P(\delta > 0) = 1$; for NG: $\delta = 0.04$, CI [0.03; 0.04], $P(\delta > 0) = 1$), while intensity shows that Syll1 is stronger than Syll3 (for GN: $\delta = 2.57$, CI [2.31; 2.84], $P(\delta > 0) = 1$; for NG: $\delta = 3.70$, CI [3.46; 3.97], $P(\delta > 0) = 1$). Thus, duration appears to enhance the second word and intensity the first word. Moreover, no difference was found across information structure conditions on

⁸For each language group, the differences among conditions in duration and intensity were tested as a function of the factors CONDITION (reference level NG), SYLLABLE (reference level SYLLABLE 1) and their interaction. As random effects, the models include random intercepts for TOKEN and SPEAKER. I used regularising weakly informative priors (Lemoine 2019) for all models (for priors specifications, see the relative RMarkdown file at <https://osf.io/9ca6m/>) and ran three sampling chains for 3000 iterations with a warm-up period of 2000 iterations for each model.

the second word. In this case, periodic energy mass is useful for overcoming the conflicting patterns, and to show that the new word is stronger than the given one not only within the NG condition, but also in GN (Section 2.3.1).

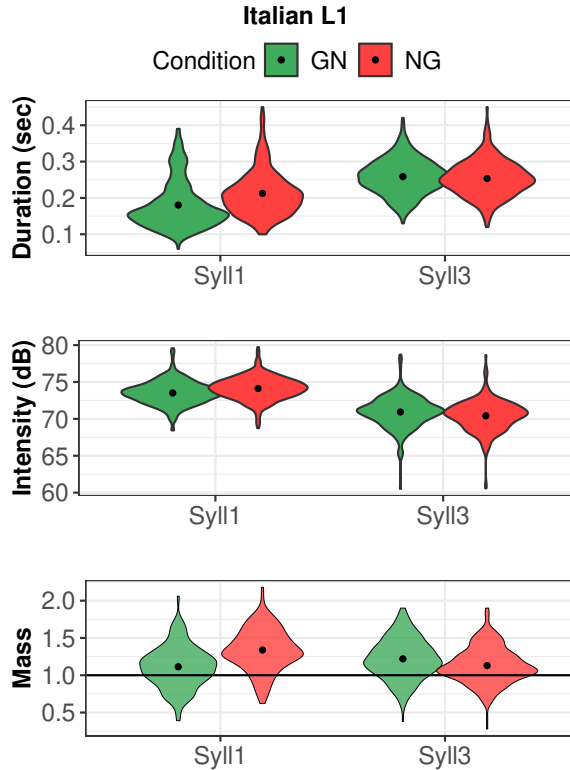


Figure 2.24: Aggregated values of duration, intensity and mass (for comparison) pooled across Italian L1 speakers. The x-axis displays Syll1, the stressed syllable of the noun, and Syll3, the stressed syllable of the adjective. Information structure conditions are colour-coded: green for given-new (GN) and red for new-given (NG).

In L1 German (Figure 2.25), intensity and duration across syllables and conditions follow far more similar trends than in L1 Italian, but these are mostly subtle modulations. Duration is slightly increased on new elements, but the trend is not robust (Syll1 in NG > Syll1 in GN: $\delta = 0.02$, CI [-0.01; 0.06], $P(\delta > 0) = 0.91$; Syll3 in NG < Syll3 in GN: $\delta = 0.01$, CI [-0.01; 0.05], $P(\delta > 0) = 0.86$). Intensity is decreased on the post-focal given element in the NG condition only (Syll3 < Syll1 in NG: $\delta = 3.83$, CI [2.83; 4.70], $P(\delta > 0) = 1$). In comparison to these results, mass seems to most clearly reveal an evident attenuation on the post-focal given element (Section 2.5).

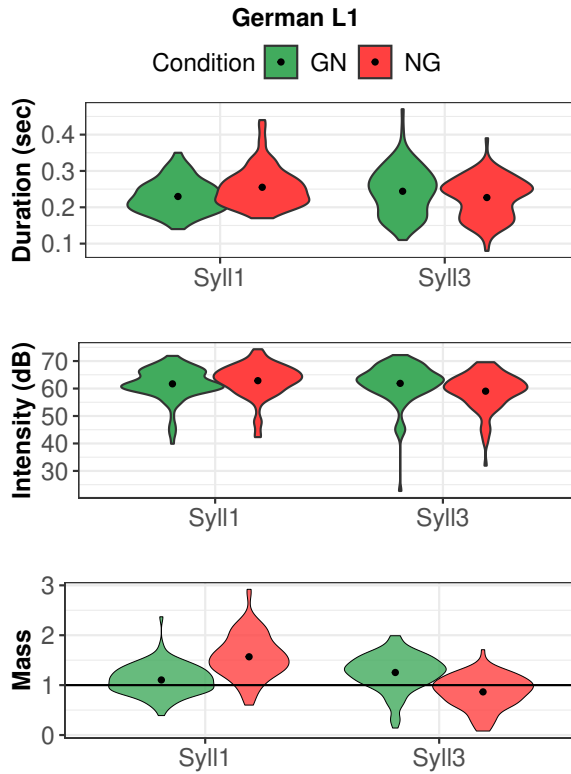


Figure 2.25: Aggregated values of duration, intensity and mass (for comparison) pooled across German L1 speakers. The x-axis displays Syll1, the stressed syllable of the adjective, and Syll3, the stressed syllable of the noun. Information structure conditions are colour-coded: green for given-new (GN) and red for new-given (NG).

In L2 German (Figure 2.26), learners use duration in a unique way that does not resemble either L1 Italian or German, in attenuating the second word of the NP in both information structure conditions by almost halving its duration as compared to the first word (Syll1 > Syll3 in GN: $\delta = 0.22$, CI [0.20; 0.24], $P(\delta > 0) = 1$; in NG: $\delta = 0.23$, CI [0.21; 0.25], $P(\delta > 0) = 1$). Intensity contributes to the reduction of post-focal material analogously to their target language, but this is not comparable to the extreme modulation of duration values, which is exploited more by learners (Syll3 < Syll1 in NG: $\delta = 1.78$, CI [1.39; 2.14], $P(\delta > 0) = 1$). These trends are well summarised also by the measure of mass (Section 2.6), which clearly highlights the peculiar behaviour of learners.

Overall, mass patterns showed differences across and within conditions more clearly, supporting the interpretation that speakers modulate prosodic strength.

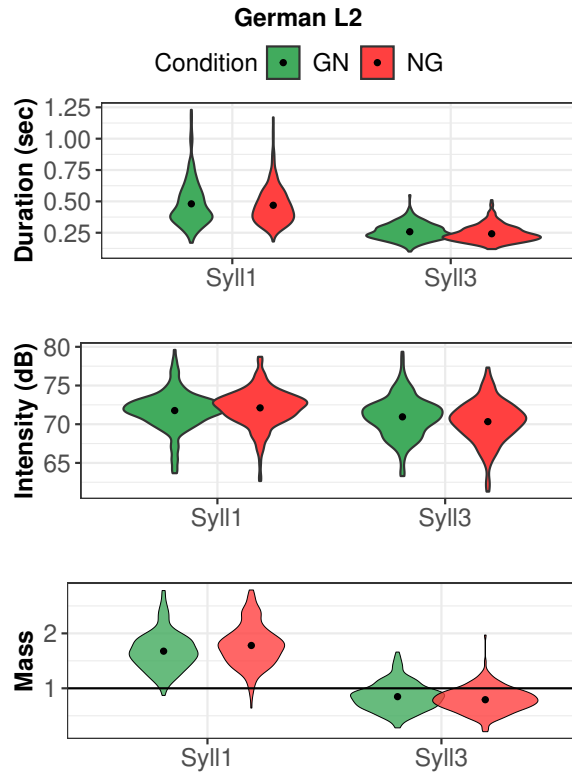


Figure 2.26: Aggregated values of duration, intensity and mass (for comparison) pooled across German L2 learners. The x-axis displays Syll1, the stressed syllable of the adjective, and Syll3, the stressed syllable of the noun. Information structure conditions are colour-coded: green for given-new (GN) and red for new-given (NG).

2.8.4 A phonological interpretation

The present investigation, based on continuous measures, does not preclude a categorical analysis in terms of pitch accent types. The results yielded by periodic-energy-based measures as well as the established measures of alignment, scaling, duration, and intensity suggest that categorical decisions are made by speakers of all groups when prosodically marking IS. Therefore, I will suggest a possible phonological description in terms of pitch accent types using a ToBI analysis (Grice et al. 2005 for German; Grice et al. 2005 for Italian) for the example noun phrases proposed in Section 2.3, Section 2.5 and Section 2.6 for the three language groups. I will concentrate exclusively on the pitch accents within the target noun phrase, and I will leave out the boundary tones since they are (for these exam-

ples) exclusively low/falling (L-L%). Table 2.3 summarises the pitch accent types found in the following example noun phrases across language groups.

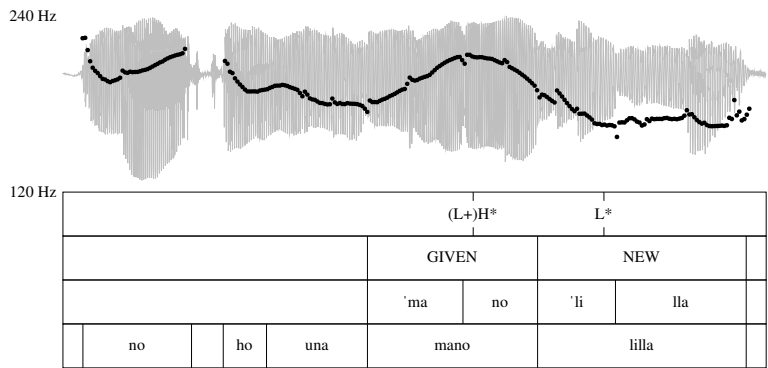
Table 2.3: Pitch accent types found in the example noun phrases.

	GN (1st word)	GN (2nd word)	NN (1st word)	NN (2nd word)	NG (1st word)	NG (2nd word)
Italian L1	(L+)H*	L*	L+H*	L*	H*+L	L*
German L1	Ø	L+H*	L+H*	(L+)H*	(L+)H*	Ø
	H*	H*	H*	(H+)!H*		
German L2	(L+)H*	L* or Ø	(L+)H*	L* or Ø	H*+L	L* or Ø

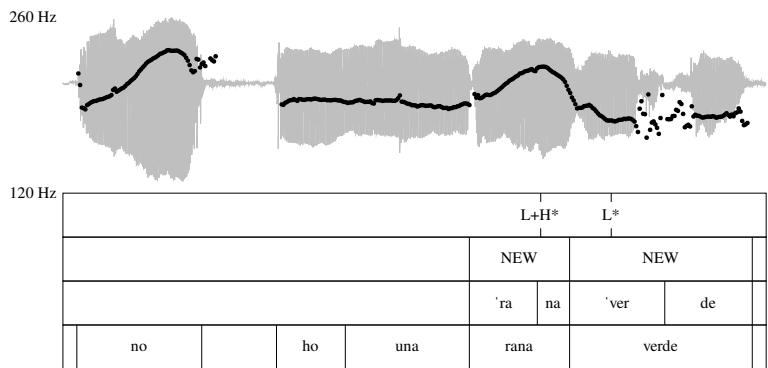
Regarding Italian, it can be concluded that, in line with previous results, prosodic marking of information status 1) is not realised through deaccentuation and 2) does not take place on the second word of the noun phrase. Indeed, in the case of the contrast between NG and NN, the acoustic properties of F0 on the first word allow for the interpretation of the information status on the second word. The difference in the acoustic properties of the first word could be prosodically represented as a difference in pitch accent type. As shown in Figure 2.27, the earlier F0 peak on the first word of NG may be described as H*+L, while the later F0 peak on the first word of GN and NN as (L+)H*. The trailing L tone is added in the former case to highlight that H is much closer to the syllable onset than in the latter case and that the fall occurs mainly on this syllable. The accent on the second word is similar in all three conditions and can be analysed as L*.

For German L1, the results are in line with the literature and show that the prosodic marking of information status is achieved by 1) the deaccentuation of post-focal given material and 2) the tendency to align an F0 peak with new or focussed elements. This alignment is generally interpreted as a H* or L+H* pitch accent. Therefore, in the case of NG, the acoustic properties of the first word of the noun phrase can be described with an (L+)H* pitch accent and the second word as deaccented. In GN and NN, there is no post-focal material, as the noun phrases are sentence-final. The typically expected realisations for GN and NN are displayed in Figure 2.28 and can be described as L+H* for the second word in the GN condition, and L+H* for the first word and (L+)!H* for the second word in the NN condition. In contrast, the hat patterns displayed in Figure 2.29 can be described as H* for the first word and H* or (H+)!H* for the second word.

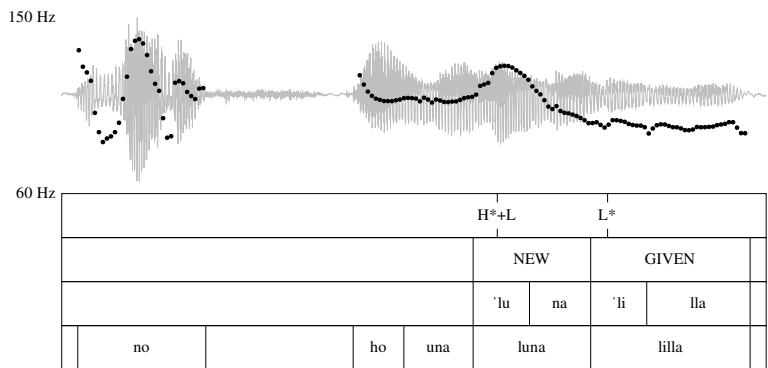
2 Prosodic marking of information status



(a) Given-New on *mano lilla* (Eng. 'lilac hand').



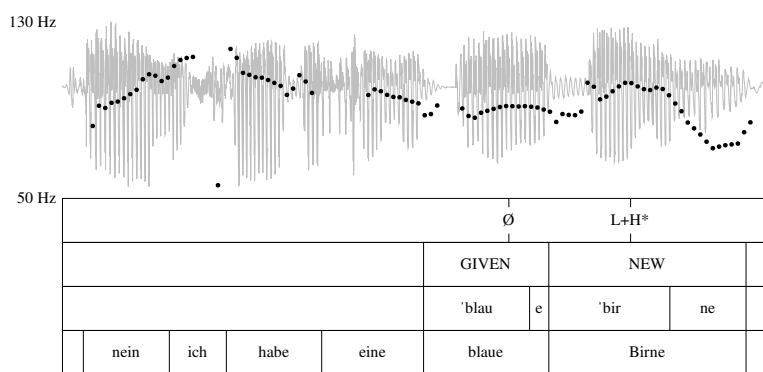
(b) New-New on *rana verde* (Eng. 'green frog').



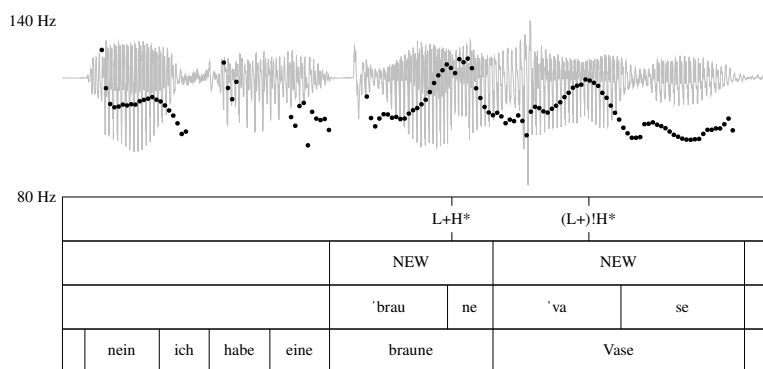
(c) New-Given on *luna lilla* (Eng. 'lilac moon').

Figure 2.27: ToBI annotation for L1 Italian. The tiers from top to bottom contain: the ToBI labels for the target noun phrases, their information status, the transcription of the stressed and unstressed syllables of the noun phrases, and the orthographic transcription of the carrier sentence.

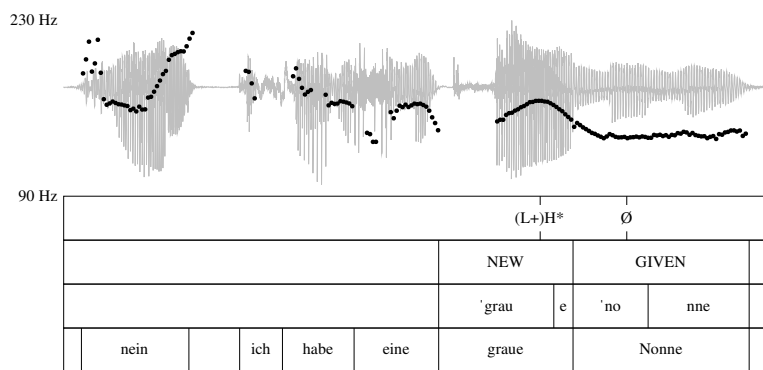
2.8 Implications for phonological analysis



(a) Given-New on *blaue Birne* (Eng. 'blue pear').



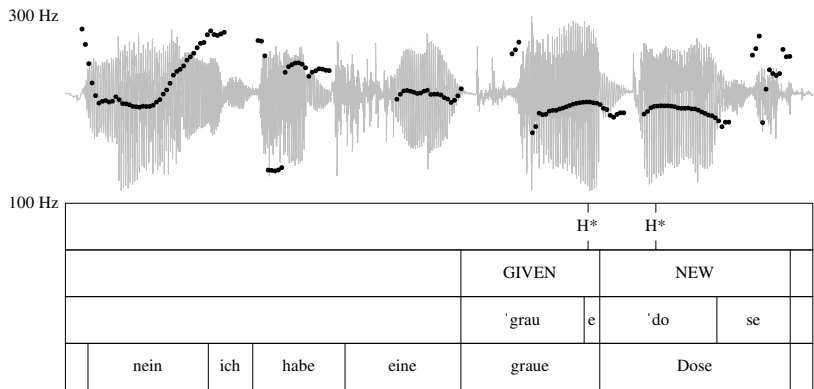
(b) New-New on *braune Vase* (Eng. 'brown vase').



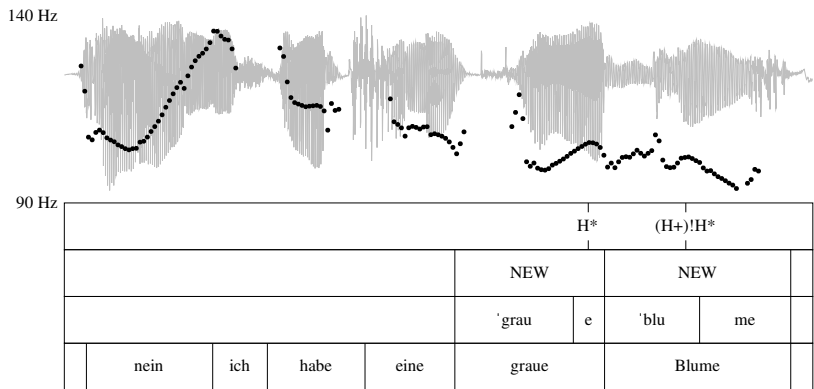
(c) New-Given on *graue Nonne* (Eng. 'grey nun').

Figure 2.28: ToBI annotation for L1 German. The tiers from top to bottom contain: the ToBI labels for the target noun phrases, their information status, the transcription of the stressed and unstressed syllables of the noun phrases, and the orthographic transcription of the carrier sentence.

2 Prosodic marking of information status



(a) Given-New on *graue Dose* (Eng. 'grey can').



(b) New-New on *graue Blume* (Eng. 'grey flower').

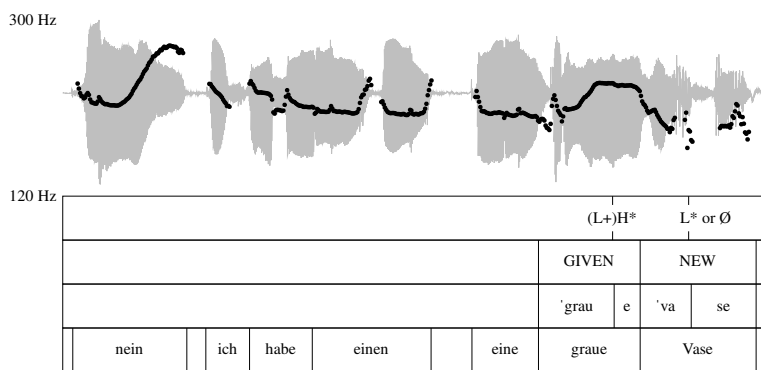
Figure 2.29: ToBI annotation for L1 German (hat pattern). The tiers from top to bottom contain: the ToBI labels for the target noun phrases, their information status, the transcription of the stressed and unstressed syllables of the noun phrases, and the orthographic transcription of the carrier sentence.

The main difference to Italian is that in native German the acoustic properties contributing to marking information status are modulated in situ (Kügler & Calhoun 2020), so that both the first and the second word of the noun phrase are affected by changes in F0 and prosodic strength based on their information status. In Italian, however, changes in the acoustic properties of only the first word contribute to distinguishing the information status of the second and last word of the noun phrase. For German L2, learners prosodically mark the information status by 1) modulating the alignment of F0 on the first word, as in their native Italian, which may be reflected in two different pitch accents, described as H*+L on the first word of NG tokens and H* or L+H* on the first word of GN and NN tokens. Moreover, 2) they do not prosodically differentiate the second word, probably due to the influence of the phonological rules of their L1 Italian. The acoustic details provide evidence for prosodic attenuation across all conditions, akin to the native German post-focal condition. For instance, in Figure 2.30, the second word could be interpreted both as L* and as deaccented. In the post-focal condition, the flat F0 on the second word could be interpreted as deaccented by a German native speaker since the more salient marking of the first new word with an earlier F0 peak can lead to the perception of lower prominence on the following given word. However, the last words sound very similar across conditions.

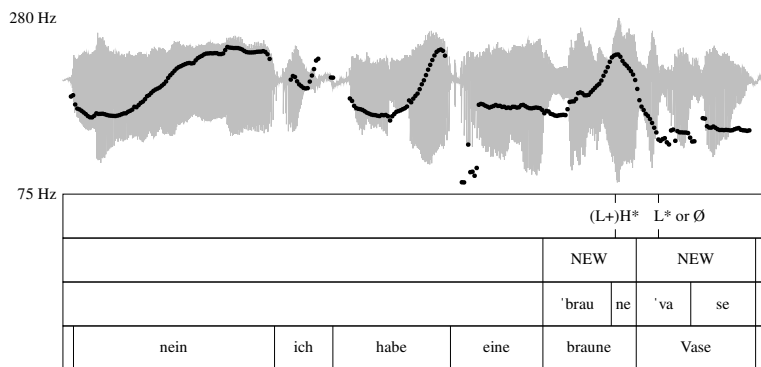
2.8.5 Summary and discussion

The results on alignment and scaling confirm the findings from the continuous measures regarding differences in F0 contours used to mark information status across all groups. This provides further evidence for the phenomena observed, independently of the method and framework used. These measures present some limitations, however, as alignment is specific to F0 configuration of peaks and other F0 shapes cannot be optimally captured (unless decisions are made as to where the peak is labelled in other F0 configurations). For this dataset, this resulted in the loss of data points, especially in L1 German. A further disadvantage is that alignment requires manual annotation of syllable boundaries, which inevitably implies a certain degree of arbitrariness. From this point of view, periodic energy measures of F0, synchrony and $\Delta F0$, have the advantage of being based on the signal itself, i.e. on periodic cycles roughly corresponding to syllables. This avoids annotator-specific decisions and permits a description of all F0 configurations in incorporating acoustic information about the underlying segmental material. Mass yielded clearer patterns than duration and intensity, as in

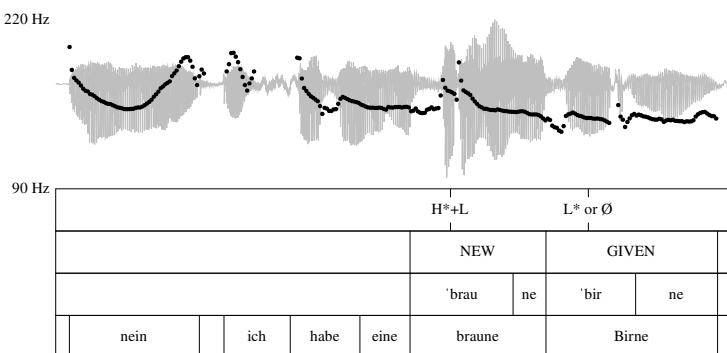
2 Prosodic marking of information status



(a) Given-New on *graue Vase* (Eng. 'grey vase').



(b) New-New on *braune Vase* (Eng. 'brown vase').



(c) New-Given on *braune Birne* (Eng. 'brown pear').

Figure 2.30: ToBI annotation for L2 German. The tiers from top to bottom contain: the ToBI labels for the target noun phrases, their information status, the transcription of the stressed and unstressed syllables of the noun phrases, and the orthographic transcription of the carrier sentence.

some cases (Italian) the latter measures even produced conflicting results. Moreover, mass can be considered more ecologically valid than duration and intensity, as these are not experienced as separate entities in actual perception.

Regarding a possible phonological interpretation, decision-making was straightforward for the two native languages, but was more complicated in the case of the interlanguage, for which I proposed a two-fold interpretation, i.e. describing the post-focal element either as L^* or as deaccented. There are several reasons for this. First, the nature of annotation itself – it is subjective, affected by expectations and the perception of meaning, as well as by the annotators' native language (Cangemi & Grice 2016). Indeed, annotators rely on F0, energy and tonal context when labelling phonological categories, reflecting their native perception. As a consequence, it is not always easy to make a decision as to which label to use, which, in turn, means that naive listeners, too, might find it hard to interpret the meaning. Moreover, it is very likely that this accentuation pattern is not consistent across all L2 noun phrases, since interlanguages are developing systems in which categories are constantly updated based on the linguistic input learners receive. Therefore, any potential category identified in L2 systems can only be taken as a snapshot of the system at that precise moment in time.

2.9 Conclusion

The present study was inspired by a series of comparative cross-linguistic studies (Avesani et al. 2013; Avesani et al. 2015; Krahmer & Swerts 2008; Swerts et al. 2002). The authors of these studies performed a categorical analysis of presence or absence of pitch accents and pitch accent types. They found that, contrary to native speakers of West Germanic languages, Italians do not mark information status prosodically within noun phrases, either in their L1 or their L2 German, which was also confirmed in perception.

The goal of the current study was to find out whether a close inspection of continuous phonetic parameters would bring to light subtle modulations for prosodically differentiating information structure, which did not emerge in previous categorical analyses. To this aim, periodic-energy-related measures were used to analyse F0 shape and prosodic strength in two-word noun phrases under different information structure conditions, i.e. given-new (GN), new-new (NN) and new-given (NG), produced by L1 Italian (Neapolitan variety) learners of German, in both their native and second language. Moreover, every attempt was made to overcome some limitations of previous studies by collecting a larger sample of data and using a more interactive elicitation method.

Results for the German native group are in line with the literature, providing evidence for deaccentuation as a marker of post-focal given information. Findings on Italian learners of German, instead, contrast with previous results. Despite reports in the literature that L1 Italians fail to deaccent, results show that the Italian participants to the current study clearly and consistently mark focused new information when given information is postfocal through different F0 peak locations, both in their L1 and in their L2, and that this difference is perceptually valid. The different contours found can be described as different accent types, but, in learners' native Italian, not by the absence of a pitch accent. In L2 German, learners prosodically attenuated the last element of a noun phrase at all proficiency levels, as in the post-focal given L1 German condition. However, attenuation is applied across all information structure conditions, meaning that prosodic strength is not used to mark different information status conditions. One possible interpretation of this result is that learners identify a reduction in prosodic strength, which can perceptually result in deaccentuation, as a salient marker of native German speech, but without identifying the relevant context. This might be due to negative interference from their L1 Italian, in which F0 peak location only (not prosodic strength) is used to discern information status contrasts within noun phrases.

One legitimate question which might arise is why these results on L1 Italian and L2 German have not emerged before. This could have two possible reasons: the methodological approach to data analysis and the elicitation method chosen, which I will now discuss.

The current study of continuous parameters revealed patterns which did not emerge previously in purely categorical analyses. Labelling phonological categories entails some degree of subjectivity due to the annotator-specific perception of meaning and expectations based on native language. As a result, different annotators can make different choices, and the individual-specific bias is even more problematic when labelling an L2. Thus, an investigation of the modulations of continuous acoustic parameters can offer a deeper understanding of linguistic phenomena by providing acoustic evidence for a categorical description. In other words, the two approaches can and should complement each other, especially when analysing complex and dynamic systems like interlanguages, where categories undergo a continuous process of restructuring based on the input and feedback that learners receive.

Another factor that may have contributed to this discrepancy with previous findings is the data collection method. Indeed, speech style and degree of spontaneity have a great influence on intonation (for the difference between read and spontaneous speech in Italian varieties and German, see De Ruiter 2015;

Grice et al. 1997). Previous studies have used an elicitation game structured in the form of alternating statements, with noun phrases containing contrastive information status categories. The production of alternating statements may not have created an engaging interaction for speakers, who may not have perceived the other player's sentences as the context for their own productions, and, instead, may have concentrated on their own list of statements. In contrast, the design of the current elicitation game was intended to create real interaction between participants, as well as engagement in the task which, together with the lack of eye-contact, may have promoted the use of prosody for conveying different pragmatic meanings, as demonstrated by perceptual results.

For Italian learners of German, however, I also found a minority of NG items realised in the same way as GN and NN items, that is, without matching prosody to the information status of the last element. This can be explained through some limitations of this elicitation method. First and most obvious, in the context of the elicited noun phrases, prosody is not necessary for the correct interpretation of the sentence, as meaning can be conveyed by the lexicon alone. Secondly, speakers may not have paid attention to the question posed by the interlocutor, since game turns are repetitive in their structure, an effect that cannot be completely avoided despite being limited by the insertion of distractors in the design. Finally, speakers may have chosen different strategies for accomplishing a task. Instead of a listener-oriented strategy, where speakers try to make sure that the interlocutor can receive and interpret the message properly, some speakers may have applied a self-oriented strategy. A prerogative to win the board game was to correctly write down all the images named by the interlocutor. As a result, some speakers may have moved their attention from the interlocutor's questions to the writing task and, consequently, produced their answers without having in mind the pragmatic context of the questions.

A further limitation of this study is that, although the interactivens of the elicitation method was improved, the type of speech investigated here cannot be described as spontaneous. More spontaneous data collected with a map task (Anderson et al. 1991) provided us with some rare, spontaneous examples of NG contours in Italian and L2 German (which can be inspected in the OSF repository at <https://osf.io/9ca6m/>). These resemble the one elicited with the semi-spontaneous board game, reassuring us that the data collected might indeed mirror non-scripted interactions. On the one hand, it is true that reliable evidence can only be drawn by systematic observation based on a large amount of data, which necessitates experimental control in the data collection process. On the other hand, future research should strive for the best possible compromise between spontaneity and systematic data collection, aiming to collect real

spontaneous conversational speech in order to increase the ecological validity of research findings.

As for possible practical applications, the results of the present study are highly relevant for language pedagogy. I found that learners do not adjust the use of prosody dependent on information status, showing that they might not be aware of this function of prosody in the target language, and fail to convey the correct meaning by prosodic means. Indeed, they tend to use marked structures (i.e. prosodic attenuation typical of post-focal given material) also in the unmarked case, differently from what one could expect according to the Markedness Differential Hypothesis (Eckman 1977), predicting more difficulty in learning marked structures (such as accentuation according to pragmatic contexts) than unmarked ones (such as accentuation in the default condition). This means that the implicit learning of prosody, i.e. without formal instructions, is not sufficient for learners to correctly acquire the target patterns, and that explicit training might help them to improve their communication skills in the L2. Some forms of prosodic training for Italian learners of German have been tested (Dahmen 2013; Missaglia 2007) and showed an improvement in learner performance in the both the segmental and suprasegmental domain. However, results on prosodic training across languages are contradictory, with some studies showing no robust (Baills et al. 2022; Suter 1976) or only minor effects (Purcell & Suter 1980). Moreover, these studies are not directly comparable in terms of native and target languages as well as methodology, making the identification of the best approach for L2 classroom setting very difficult (e.g., hand gestures in Baills et al. 2022; imitation and repetition in Nicora et al. 2018; contour visualisation using Praat in Smorenburg et al. 2015; contour visualization using stylised contours in Niebuhr et al. 2017; computer- or robot-assisted techniques in Bissiri 2008; Fischer et al. 2021). Therefore, more research on efficient and practicable pedagogical techniques for teaching prosody in L2 classrooms is still necessary.

The present study has provided useful theoretical groundwork for the development of such pedagogical tools by individuating critical aspects of L2 prosody acquisition and exemplifying the prosodic strategy for marking information status in learners' native and target languages. The analysis of learners' native and target languages provides teachers with empirical knowledge on the pragmatic use of prosody in both phonological systems contributing to learners' interlanguage, instead of a (complete) reliance on native speaker intuitions (Derwing & Munro 2015). The analysis of learners' difficulties in acquiring the target prosodic features highlights the aspects of prosodic competence which should be explicitly thematised in teaching and be embedded in pedagogical tools to ensure their successful acquisition.

3 Turn-taking fluency in L2 interactions

This chapter includes a study dedicated to a specific ability of L2 interactional fluency (Peltonen 2024, 2020), that is turn-taking in dyadic interactions.¹

I problematise the discrepancy between the theory of CEFR (Council of Europe 2020) that describes learners as “social agents”, emphasising the interactional aspect of SLA, and the practice of most assessment realities, focused on the quantification of grammar and lexical competence, while neglecting interactional aspects. An obstacle to assessing learners in interactions is the complexity of the communicative phenomena involved in this process and the shared responsibility of participants in the co-construction of meaning.

With this exploratory study, I try to break down the complexity of interaction by starting from its very foundation: the management of the conversational floor. I test a method for visualisation and quantification of turn-taking fluency across L2 proficiency levels by extracting reliable and testable metrics. This method could represent a valid starting point to build upon and develop a more complete tool for a quantifiable assessment of interactional competence.

3.1 Background

3.1.1 Assessing interactional competence

The assessment of learner proficiency often tends to focus mainly on grammar and the lexicon, neglecting interactional aspects. Many language test formats only involve a written form, such as the cloze format, where some words of a text are replaced with gaps to be filled in by learners. These kinds of test are often used with the explanation that they show correlations to all receptive and productive abilities: reading, listening, writing and speaking (Council of Europe 2001). Nevertheless, they mainly put to test grammar and vocabulary knowledge, leaving out the full range of pragmatic and strategic resources required for oral interaction.

¹The analysis provided in this chapter is an extended and enriched version of previous work published in Sbranna et al. (2020).

3 *Turn-taking fluency in L2 interactions*

Bachman & Palmer (1996), in “Language Testing Practice”, mention interactivity as one of the fundamental characteristics a good quality language proficiency test should have: reliability, construct validity, authenticity and interactivity. Reliability refers to the consistency of the measurement, and thereby of the results given by the test. Construct validity indicates the possibility of interpreting the score of the test as a valid indication of global language proficiency. Authenticity defines to what extent the task given to learners in test circumstances corresponds to real-life tasks they would perform using the L2. Finally, interactivity refers to the degree of involvement of learners’ different abilities in accomplishing the task, i.e. the extent to which a test involves various learners’ skills, which include general language knowledge, metacognitive strategies and strategic competence for planning and dealing with unexpected difficulties, topical knowledge, and affective schemata, which refers to learners’ emotional response to the task. According to the authors, tasks with a high level of interactivity are role-playing and long conversations, as they require learners to draw on all these abilities.

Some possible reasons for neglecting highly interactive tasks in language proficiency testing may be practical. He & Young (1998) point out that having learners interviewed by native or highly proficient speakers can create certain difficulties. First, such interviewers have to be available; secondly, the interviews need to be carried out for a reasonable length of time to allow the interviewer to elicit enough linguistic data from the learner so that these data can be considered representative of the learner’s global knowledge. Hence, such testing would require more assessors, be time-consuming and would consequently be more expensive than a test format such as the cloze test, which can optimise time for testing and correction. Even when an assessment of the L2 speaking ability is conducted, the quantification of the skills involved in interaction may turn out to be extremely time-consuming and complex to synthesise. As a result, interactional competence is often subject only to a qualitative evaluation based on illustrative scales,² the interpretation of which may include a certain degree of subjectivity. Another problematic aspect of L2 oral proficiency assessment in non-experimental environments is the method used, i.e. oral proficiency interviews where a native speaker tries to elicit linguistic information from learners

²An example from the illustrative scale for overall spoken interaction (C2 level): “I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely. If I do have a problem I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it.” (Piccardo et al. 2018: 168).

using a script representative of real-life language settings. Although these interviews try to simulate ordinary conversation, they are limited by various constraints that can affect learners' oral performance (He & Young 1998): interviews take place in an institutional setting; speech activities are predetermined; and participants have different statuses, two different L1s and cultural backgrounds, as well as two different proficiency levels of the language used during the interviews. Moreover, being an interaction among a tester and a learner, the attention is totally on the learner's production, their fluency and accuracy, and not on the interactional patterns they are able to create, maintain and manage during the exchange with the tester. Consequently, 1) interviews risk not providing a snapshot of learners' skills that would be representative of a more spontaneous and less formal style, such as peer conversations, and 2) risk focussing on learner fluency and accuracy, instead of on more specifically interactional abilities.

In line with this tendency in non-experimental environments, research has mainly focused on the measurement of *individual* fluency – how fluent learner speech is *within* turns – rather than *interactional* fluency – how fluent the conversation is *across* interlocutors' turns – (Peltonen 2017) with the aim of proposing quantification methods for L2 speaking abilities. Fluency is consistently mentioned as a fundamental component of learners' oral performance in various assessment traditions, and its correlation with general L2 proficiency has been demonstrated by several studies (De Jong et al. 2013, 2015; Segalowitz & Freed 2004). Therefore, the following paragraphs present a brief review of theories and findings on fluency as a measure of L2 speaking proficiency.

3.1.2 Fluency in L2 studies

Fluency has been identified as one of the main aspects that ensure the success of a speaking performance (De Jong 2016). One of its first definitions can be traced back to Fillmore (1979). He defines fluency as a measure of how well a language is spoken, in other words, the skill of using L2 knowledge efficiently, and enumerates four dimensions of fluency, including both quantitative and qualitative aspects: the ability to speak at length with few breaks; the ability to speak in a coherent, reasoned, and semantically dense way; the ability to talk appropriately according to context; and the ability to be creative and imaginative in speech production.

Lennon (1990) distinguishes between a broad and a narrow definition of fluency. In the broad sense, fluency encompasses overall language proficiency, including grammatical accuracy and vocabulary knowledge. In contrast, fluent speech in its narrow sense focuses specifically on the ease and automaticity of

3 *Turn-taking fluency in L2 interactions*

speech production and is defined as “unimpeded by silent pauses and hesitations, filled pauses (“ers” and “erms”), self-corrections, repetitions, false starts, and the like” (Lennon 1990: 390). Tavakoli & Hunter (2018) built upon these concepts by exploring how teachers perceive and teach fluency. Specifically, they argue that many teachers interpret fluency primarily in its broad sense, while remaining largely unfamiliar with its narrow sense.

Starkweather (1987) instead suggests four dimensions of fluency mainly related to physical aspects of speech: continuity, rate, rhythm, and effort. In other words, fluent speech should present few discontinuities, have a regular rhythm and a fast rate, and not require too much cognitive and physical effort (Zmarich 2017).

In his model of fluency, Logan (2015) adds two additional dimensions: naturalness, i.e. how much speech resembles that uttered by a typical speaker with regard to continuity, rate, rhythm and effort; and stability, i.e. how similar a speaker’s performances are over time if subject to repeated measurements.

Segalowitz (2010) focuses on L2 fluency from a dynamical systems perspective. He argues that fluency is strongly linked to the social context in which the speech performance takes place and distinguishes three aspects: utterance fluency, cognitive fluency, and perceived fluency. L2 utterance fluency refers to the fluidity observable in a speech sample and quantifiable by temporal measures, among which the author mentions syllable rate, duration and rate of hesitations, filled and silent pauses, breakdown fluency (pausing phenomena) and repair fluency (false starts, corrections, repetitions). Indeed, most studies calculating temporal measures for fluency follow the classification in the sub-components breakdown, repair, and speed fluency (Tavakoli & Skehan 2005; Huensch & Tracy-Ventura 2017; Tavakoli et al. 2020; Lahmann et al. 2017). L2 cognitive fluency refers to the fluidity of the cognitive processes underlying speech production, such as processing skills (declarative and procedural knowledge), efficiency and speed of semantic retrieval, and cognitive load in working memory. Some measures of cognitive fluency have been found to correlate with L2 proficiency, e.g. reaction time and switch cost among competing tasks. Reaction time and its coefficient of variability have been used to operationalise the efficiency of semantic retrieval (Segalowitz & Freed 2004), while switch cost has been used as an indicator of linguistic attention, which refers to attention shifting guided by connections among grammatical elements within utterances (Duncan et al. 2014).

Such a systemic understanding of fluency is also assumed by Kormos’ (2006) psycholinguistic model, in which different cognitive processes underlie the three above-mentioned sub-components of fluency. In particular, breakdown measures are related to learner effort. For instance, final-clause pauses reflect a learner’s

conceptualization and planning of the message, and mid-clause pauses represent the time taken by learners to encode and formulate linguistic information, while repair measures signal the monitoring of the speech output and consequently the amount of attention required for speaking the L2. Finally, speed-related measures provide information on the degree of automatization in all the above processes.

Another important aspect that should be taken into account is the fact that both utterance and cognitive fluency are specific to each person. Still, individual variability in an L1 can only partially explain individual variability in an L2 (De Jong et al. 2013) since disfluency is also characterised by L2-specific features, such as a higher cognitive load. Therefore, it may be good scientific practice to consider L1 fluency measures as a baseline (as in De Jong et al. 2015; Saito et al. 2018) to get a clearer picture of L2-specific fluency measures by partialling out the variables that are not specifically related to L2 disfluency phenomena (Segalowitz 2010).

Finally, perceived L2 fluency indicates subjective listeners' ratings on how fluent a speaker is. One disadvantage is that, being subjective, perceived fluency is only moderately informative about utterance fluency and cannot explain all the variance in objective measures. However, it is helpful to gain an understanding of which cues are relevant to native listeners when judging L2 speech fluency in relation to L2 proficiency. Moreover, a listener's judgment of their interlocutor's fluency can affect the interaction and influence both speakers' fluency.

3.1.3 Fluency measures and operationalisations

Research on L2 fluency has focused on individuating which objective measures can best explain L2 fluency judgments and has mainly concentrated on temporal features. A categorisation of aspects of fluency comparable to the already-mentioned, more recent triad of "breakdown, repair and speed" (Tavakoli & Skehan 2005) was already proposed in one pioneering study. Riggensbach (1991) classified the features which can characterise a judgement as fluent or non-fluent in non-native speech into hesitation and repair phenomena as well as rate and amount of speech. The study also includes an analysis of interactive features contributing to the turn-taking alternation, such as overlaps, pauses between turns, and collaborative completions. Hesitation phenomena and speech rate were found to be significantly correlated with ratings of L2 fluency, with hesitation placement and the resulting discourse chunking playing a central role. In contrast, results related to repair phenomena appeared to be less clear, probably due to the small amount of data. The same holds for interactive features, which showed high variability due to the idiosyncratic nature of interactions, which

3 *Turn-taking fluency in L2 interactions*

vary according to many linguistic and non-linguistic factors. Furthermore, this pragmatic-oriented analysis is reported to be extremely time-consuming and, indeed, studies on turn-taking fluency are relatively scarce.

Later on, many studies confirmed that perceived fluency by both native (Bosker et al. 2013; Derwing et al. 2004; Kormos & Dénes 2004; Préfontaine et al. 2016; Rossiter 2009; Saito et al. 2018; Suzuki & Kormos 2020) and L2 listeners (Magne et al. 2019) is closely related to speed of delivery and pausing phenomena.

As reported in Suzuki & Kormos (2020), a recent approach in research on fluency differentiates three independent dimensions of breakdown fluency – pause frequency, duration, and location – and all of them have been demonstrated to independently contribute to fluency. Moreover, concerning pause location, mid-clause pauses have been found to have a more distinctive role than clause-final pauses (Magne et al. 2019; Saito et al. 2018; Suzuki & Kormos 2020). One possible reason, following Kormos' (2006) model, would be that mid-clause pauses, being associated with the time required for linguistic encoding, are more representative of proficiency than clause-final pauses, which are associated with content planning.

However, there are several differences in methodology across these studies, in particular regarding rating methods, the task used for data collection, and the operationalisation of measures. Table 3.1–3.4³ summarise the most frequently listed temporal measures in literature reviews on fluency, as well as some interactional measures. Moreover, they differ considerably in sample size.

In particular, Derwing et al. (2004) argue that tasks used in experimental and assessment settings can have a considerable impact on learners' L2 fluency, depending on the amount of freedom accorded to the participants. For example, picture narrative and description impose a given range of lexical and syntactic structures, while a monologue or a conversation with free choice of topic allow learners to have much more control on the content and the expressions used to deliver it. For this reason, a wider range of speaking tasks should be employed, in particular more open and interactive ones, which are more representative of our daily use of spoken language. Furthermore, especially in experimental settings, assuming L1 fluency measures as a baseline for each speaker's L2 fluency can

³Measures marked by a star have been found to be significant predictors of L2 learner fluency. Notably, in a more recent framework by Tavakoli et al. (2020), articulation rate is identified as the principal measure of the speed dimension, while measures incorporating information about pauses alongside production speed are regarded as composite measures, combining dimensions of speed and breakdown.

Table 3.1: The most frequently used temporal and interactional measures in research on L2 fluency: Speed of speech measures.

	Formula	Reference
Speech rate	number of syllables / total time	Kormos (2006) in: De Jong (2016)*
Pruned speech rate	number of syllables - number of disfluent syllables / total time	De Jong (2016)
Phonation time ratio	speaking time / total time	Kormos (2006) in: De Jong (2016)*
Articulation rate	number of syllables / speaking time	Kormos (2006) in: De Jong (2016)
Mean length of syllables	speaking time / number of syllables	De Jong (2016)*
Mean length of run	number of silent pauses / number of syllables	Kormos (2006)*

Table 3.2: The most frequently used temporal and interactional measures in research on L2 fluency: Breakdown fluency measures

	Formula	Reference
<i>Frequency</i>		
Number of pauses	number of pauses / total time or speaking time	Saito et al. (2018) for references*
Number of silent pauses	number of silent pauses / total time or speaking time	Kormos (2006) in: De Jong (2016)*
Mean length of utterance	total speaking time / number of utterances	De Jong (2016)
Number of filled pauses	number of filled pauses / total time or speaking time	Kormos (2006) in: De Jong (2016)
<i>Duration</i>		
Duration of silent pauses	pausing time / number of silent pauses	Kormos (2006) in: De Jong (2016)
<i>Location</i>		
Number of clause-medial pauses	number of clause-medial pauses / total time or speaking time	Saito et al. (2018)*
Number of clause-final pauses	number of clause-final pauses / total time or speaking time	Saito et al. (2018)*

3 Turn-taking fluency in L2 interactions

Table 3.3: The most frequently used temporal and interactional measures in research on L2 fluency: Repair fluency measures

	Formula	Reference
Number of repetitions	number of repetitions / total time or speaking time	De Jong (2016)*
Number of repairs	number of corrections and restarts / total time or speaking time	De Jong (2016)*

Table 3.4: The most frequently used temporal and interactional measures in research on L2 fluency: Turn-taking fluency measures

	Formula	Reference
Pause	within speaker silence	Edlund et al. (2010)
Gap	between speakers silence	Edlund et al. (2010)
Overlap	turn-changing and non-turn-changing	Edlund et al. (2010)
Backchannel	not taking the turn	Riggenbach (1991)
Collaborative completion	attempt to complete a sentence or a phrase of the other	Riggenbach (1991); Peltonen (2017)
Other-repetitions	repetition of part of the turn of the other	Peltonen (2017)

help explain some idiosyncratic differences by controlling for non-linguistic factors possibly affecting learner performance (Segalowitz 2016), such as contextual factors (e.g. attitude towards the task and the interlocutor), but also participant-specific ones (e.g. personality and motivation).

Finally, it should be noted that the static perspective provided by averaged measures of learners' performance offers only a limited view. Fluency, instead, is a dynamic phenomenon that fluctuates both between speakers (inter-speaker) and within the same speaker (intra-speaker) over time. A long tradition of research has focussed on how conceptual planning impacts fluency, providing evidence for the alternation between fluent and disfluent temporal cycles during oral performance (Henderson et al. 1966; Goldman-Eisler 1967; Butterworth 1975; Beattie 1980). In this context, low-fluency sequences are believed to correspond to the conceptualisation phase of speech production, while high-fluency sequences align with the less cognitively demanding phases of formulation and articulation. Currently, there is limited research on the dynamics of L2 fluency (De Jong 2023) probably due to the fact that it is very time-consuming (Roberts & Kirsner 2000). I contribute to the field by introducing a visualisation method to explore

these dynamics, even though fluency dynamics are not the primary focus of this research.

3.1.4 Interactional fluency visualisation methods

Most of the studies mentioned in the previous paragraphs have focused mainly on the concept of fluency as an individual phenomenon (with the exceptions of Riggensbach 1991; Tavakoli 2016; Peltonen 2017, while Peltonen 2022 and Sato 2014 also considered interactional fluency, but from an assessment perspective). Mostly using monologic tasks for data collection, their main concern was to define if and to what extent a speaker is fluent, the perception from a listener's perspective, and implications for L2 assessment (see also more recent work by Götz 2013; Kahng 2014; Peltonen & Lintunen 2016).

Nevertheless, researchers agree that the circumstances in which learner-spoken performances take place are fundamental for making assumptions about the performance itself. Therefore, the measures above only provide incomplete information about L2 fluency if considered out of context. This is especially true in the case of interactions, which are collaborative in nature. For example, it has been suggested that speed of delivery and pausing behaviour are accommodated to the interlocutor during the interaction (Kousidis & Dorran 2009), so that a jointly achieved harmonisation of tempo occurs. This phenomenon has been depicted through the metaphor of interactional “flow” (McCarthy 2009). Moreover, being ruled by the turn-taking system (Sacks et al. 1974), interactions more thoroughly put to the test automaticity in L2 speech. Indeed, turn-boundaries (also called *transition relevance places* – TRP – in conversational analysis) are the places in which smooth or disfluent transition of turns can take place and which require the interlocutor to appropriately anticipate the end of their turn to be able to quickly react (Bögels & Torreira 2015; Levinson 2016). For these reasons, judgements of fluency based on a single speaker and ignoring the interlocutor's contribution to the conversation would lack the interactive perspective and important information about learners' abilities to co-create fluency, keeping in mind that interaction is described as a co-creation process in the CEFR (Piccardo et al. 2018: 81).

In the CEFR, both dimensions of individual and interactional fluency have their own descriptors in the assessment scale (Council of Europe 2001: 28–29), with fluency including aspects of individual fluency (such as speed, pausing, and repair fluency) and interaction encompassing aspects of interactional competence, “thus also containing elements that could be considered indicators of interactional fluency” (Peltonen 2017: 30). In particular, turn-taking management

3 *Turn-taking fluency in L2 interactions*

is the only ability which is exclusively mentioned under spoken interaction, whereas many fluency-related abilities are shared across communicative activities. Turn-taking appears to be the distinguishing ability of interactional competence, hence, a good starting point when examining L2 interactions.

L2 interactions have been approached from several perspectives: using conversational analysis to explore learners' interactional practices (Pekarek Doehler & Berger 2015, 2018) and, in L2 assessment studies, examining interactional cues that contribute to create cohesion (e.g., feedback in Galaczi 2014; May et al. 2020, and collaborative completions and other-repetitions in Peltonen 2017). However, very little quantitative research has been conducted on the timing of turn-taking in L2 (Sørensen et al. 2021).

Studies focusing on conversational speech rhythm have developed similar methods for capturing the speech activity performed in dyadic exchanges. Visualisation methods of this kind were pioneered by Chapple (1939) in the field of anthropology and applied later in several domains and to different research objectives in the field of speech studies (for a review see Cangemi et al. 2023). This visualisation method has been used to visualise the timing organisation of the interaction, categorised into classes of activity. The most basic version would be to represent speech vs. silence for each of the participant to the interaction, but other classes of activities can be flexibly introduced based on the specific research question. The horizontal axis of the plot displays a time window of one minute, whereas the vertical axis reproduces time passing throughout the interaction. The speaking activity is then represented on this graphic scaffold by colour-coded bars, whose length represents the duration of each speaker's turn. When the two different bars are on top of each other, speakers speak at the same time, causing an overlap, whilst when bars end and a white space follows, speakers are silent. The colour coding can be then used to identify other classes of activity.

Some examples of application include speech activity patterns in telephone conversations (Campbell 2007), the interplay of laughing and speaking (Trouvain & Truong 2013), human-machine interaction (Gilmartin et al. 2018), psychiatric interviews (i.e. the effects of disease on interaction patterns in Cangemi et al. 2023), cross-linguistic comparison of turn-taking patterns (Dingemanse & Liesenfeld 2022) and even multimodal communication (Rühlemann & Ptak 2023; Spaniol et al. 2023).

Despite being content-free, this kind of chart displaying the duration and timing of specific classes of interest has three main benefits. First, it serves as an “eye-opener” (Trouvain & Truong 2013: 4) helping researchers to evaluate their intuitions by means of visual exploration and comparison of data through a close

and analytical reading of speech activities. Secondly, in the case of speech activity, the annotation can be performed largely in an automatic way and manual verification, if required, consists in excluding intervals containing vegetative vocal activities (e.g. coughing) and undesired noises. This enables a quick analysis of large amounts of interactional data. Finally, the data extracted to create the plot provide the material for quantitative analysis and statistical testing.

In the present study, this type of visualisation is tested on L2 interactional data to assess its representative power and usefulness in capturing differences in oral interaction management by learners with different levels of L2 proficiency. Even though some hypothesis testing will be conducted for demonstrative purposes only (extending a previous investigation carried out in Sbranna et al. 2020), due to its preliminary nature, the main goal of this study is to explore the potential of this visualisation and quantification method when applied to L2 data. Specifically, the aim is to evaluate whether these tools can help fill the gap in quantification methods for L2 interactional competence and serve as possible groundwork for developing a more complete instrument for a standardised assessment of L2 oral performance.

3.2 Method

3.2.1 Corpus

The corpus consists of thirty-nine task-based dialogues by forty Italian learners of German, nineteen performed in their native language⁴ – Italian, Neapolitan variety – and twenty in L2 German categorized at different proficiency levels ranging from A2 to C1 on the CEFR scale. For the purpose of hypothesis testing, learners had been recategorised into two homogeneous groups based on their proficiency levels, i.e. beginner (from A1 to B1 levels) and advanced learners (from B2 to C2 levels). However, some qualitative considerations on the development of their conversational patterns across their actual CEFR proficiency levels will also be accounted for and discussed.⁵

The corpus also includes nine dialogues performed by eighteen German native speakers, which will enable an exploratory cross-linguistic comparison of native conversational schemata, i.e. across Italian and German as L1s, but, crucially, not as a target for L2 learners.

⁴The file for dyad ME corresponding to the dialogue in Italian language was found to be damaged and could not be analysed.

⁵All details on participants, data collection, and learner proficiency levels are presented in detail in Section 1.3.

3 Turn-taking fluency in L2 interactions

Following the suggestion of Segalowitz (2016), learner L1 will be used as a baseline against which to assess L2 interactional patterns. This represents a relatively novel approach in fluency research, which has so far primarily applied Segalowitz's idea to monologic data. As mentioned in Section 1.3.3, it is implausible to suppose that the learners' way of interacting in L2 German would be similar to an L1 German speaker, since at the moment of the recording, participants were living in Italy, studying in Italy and talking to an Italian interlocutor with whom they shared the same L1 and culture. There is not enough foreign exposure to and experience with a German native conversational style which could favour a possible native target, as in L2 classrooms, conversational skills are mostly practiced among learners themselves. Moreover, using the L1 as a baseline was informative in a previous pilot study (Sbranna et al. 2020), which showed that with increasing proficiency, the interaction in the L2 approached the same interactional pattern learners used in their own L1.

3.2.2 Elicitation method

To investigate interactional competence in SLA in classroom settings (discussed is in this and the following Chapter), I collected data using a task, inspired by the Task-Based Language Teaching approach (TBLT; Long 1985, 2015; see also Gass & Mackey 2014) which is increasingly being applied in L2 classrooms. This approach is rooted in communicative language teaching principles and consists in engaging learners in meaningful and authentic tasks that require the use of the target language and are designed to be relevant to learners' needs and interests. TBLT views language as a tool for accomplishing communicative goals rather than as a set of grammar rules and vocabulary items. Thus, learners are encouraged to learn by using the language to accomplish tasks rather than simply practice isolated language components.

For this reason, semi-spontaneous speech data were elicited using the Map Task (Anderson et al. 1991; see Grice & Savino 2003 for set up, map layout and instructions), which matches the goal-oriented cooperation task mentioned in the CEFR, i.e. (Piccardo et al. 2018: 88).

In this task, participants are provided with two maps (Figure 3.1 for Italian and Figure 3.2 for German); one speaker receives a map with a route drawn across landmarks – the instruction giver – and has to describe the route to the other participant – the instruction follower – whose map only features landmarks. The goal is to co-operate so that the instruction follower can reproduce the route on their map thanks to the instructions given by the partner. Some landmarks are

different across maps, but participants only discover this during the task, which creates unexpected problem-solving situations.

To carry out the task, participants sat opposite each other and eye-contact was prevented using an opaque dividing panel, in order to maximise the use of the verbal channel for signalling turn-taking and providing feedback (discussed in Chapter 4).

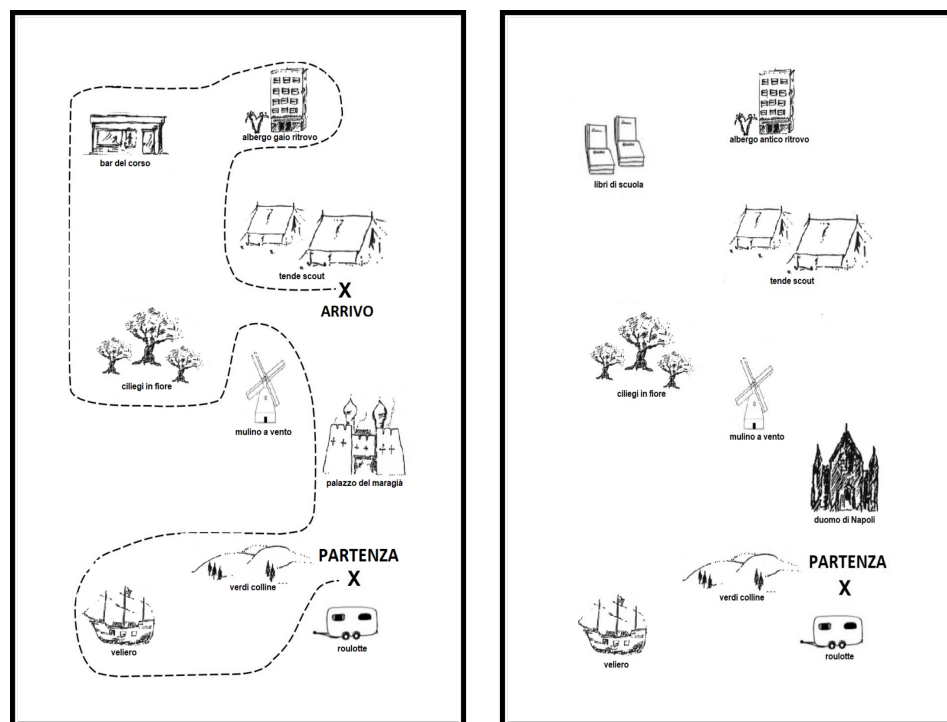


Figure 3.1: Italian version of Map Task

This task was chosen for two reasons. First, it can be performed at every proficiency level, since learners should address the topic of grammar and vocabulary knowledge related to road indications at a beginner level according to the CEFR. In addition, it presents a fair degree of openness thanks to the unforeseen unmatched landmarks, which increase the degree of spontaneity in the interaction.

Native dyads performed the task once. Learners repeated it in both their native and second language and kept the same role (either instruction giver, or follower) across the two languages to prevent cross-language differences from being attributable to factors regarding their role in the task.

3 Turn-taking fluency in L2 interactions

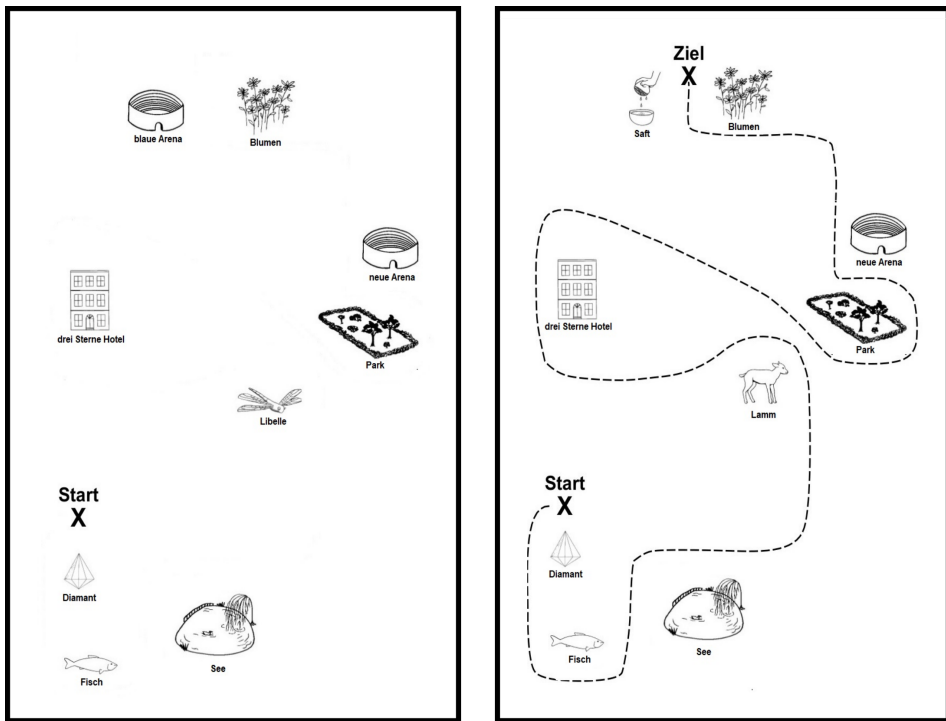


Figure 3.2: German version of Map Task

3.2.3 Metrics

As a proxy for interactional competence, I explored the degree of fluency of the interaction co-created by participants through the turn-taking system. The interactional flow was operationalised by quantifying the percentage of time of five classes of conversational activities: 1) and 2) how long each of the two speakers (giver and follower) takes the floor, 3) how long participants' turns overlap, 4) the amount of backchannels not initiating a turn and 5) the amount of total silence in the conversation. I distinguish backchannels which are not turn-initial from overlap since these tend to overlap with the interlocutor's speech, but do not represent a genuine overlap between turns (i.e. they do not represent the intention to take the floor and, instead, fulfil a function supporting the interlocutor's speech). Nonetheless, I will not focus on backchannels in this study, since the following separate study in Chapter 4 is entirely dedicated to a comprehensive analysis of this conversational phenomenon. These metrics are extracted by a Praat script sampling the classes of conversational activities at a regular time

interval of 0.1 seconds across the whole dialogue duration, and thus, dialogue duration is quantified as the total amount of time samples extracted using the Praat script.

3.2.4 Procedure for tools generation

The annotation and extraction procedure is composed of three steps. After having extracted each of the two channels from the stereo recordings, a first step consisted of the automatic labelling of interpausal units in Praat using the function of silent interval detection with a minimum duration of 200 ms (Gleitman et al. 2007; Griffin & Bock 2000; Levinson & Torreira 2015; Schnur et al. 2006; Wesseling & van Son 2005). Secondly, boundaries were manually checked and corrected to make sure that interpausal units were correctly identified, since some voiceless consonants were automatically labelled as silence. Finally, the annotation text files (Praat TextGrid files) related to the two audio channels of each dialogue were used as input files for a Praat script (Cangemi et al. 2023), which generated a figure depicting each speaker's contribution to the interaction as it develops over time (Figures 3.3, 3.4, 3.5, 3.6).

In this conversation chart, each horizontal bar corresponds to an interpausal unit uttered by one of the two speakers involved in the task. Speakers are colour-coded according to their role in the dialogue: red for the instruction giver and blue for the instruction follower. Whenever the two speakers overlap, the differently coloured bars are on top of each other. Backchannels are depicted in green, to distinguish them from actual turn overlap. Time unfolds from top to bottom – minutes –, and from left to right – seconds – so that the interaction can be followed as on a written page in a left-to-right writing system.

Figures 3.3 and 3.4 display a low-proficiency dyad performing the task in their L1 and L2, respectively. The first striking difference is the total length; in the L2, these speakers need roughly 150% of the time they need in L1 to conclude the task. Moreover, differently from the smooth L1 pattern, the flow of interaction in L2 appears much more fragmented, with shorter turns and more frequent and longer pauses, especially during the turn of an individual speaker. This observation is in line with the systemic perspective of fluency mentioned in the background (Kormos 2006), according to which utterance fluency measures mirror learners' cognitive fluency. Indeed, in the case of this dyad with a low proficiency of German (A1 level, beginner), it is not surprising to find lengthened within-speaker pauses in the L2 as compared to their interactional behaviour in the L1, reflecting greater cognitive effort.

3 Turn-taking fluency in L2 interactions

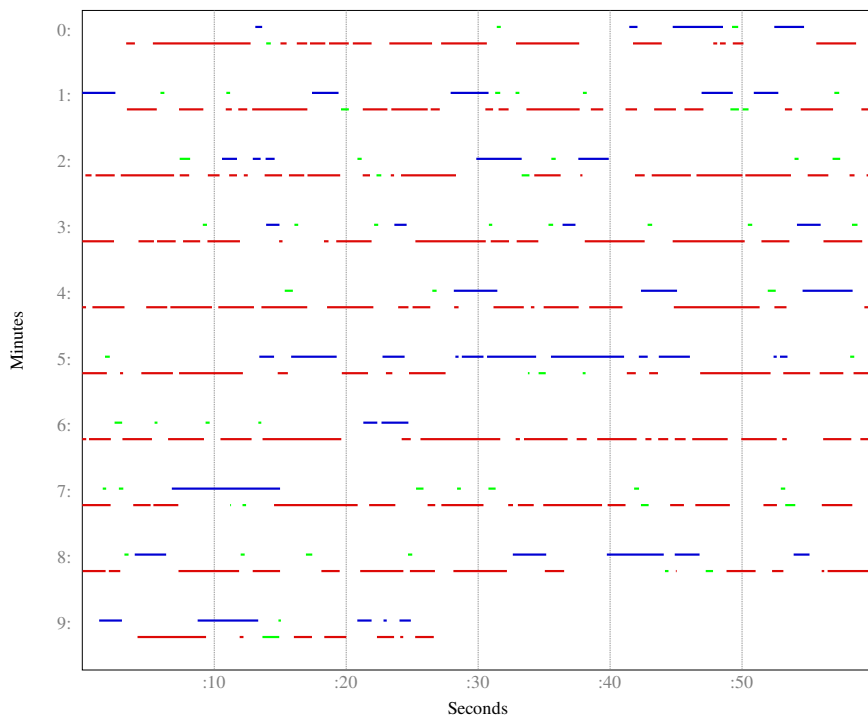


Figure 3.3: Visualisation of the interactional flow in L1 Italian (dyad with low L2 proficiency). Color code identifies speakers according to their role in the Map Task: red – instruction giver, blue – instruction follower. Green bars mark backchannels.

Figures 3.5 and 3.6 show the interactional patterns of a dyad with high proficiency of German (C1 level, advanced). In this case, it is difficult to identify at first sight which dialogue was carried out in the foreign language since the two interactional patterns look very similar. It can still be noticed that turns are slightly more fragmented in the L2, especially for the instruction follower, and that two highly proficient speakers also need a little more time to complete the task in their L2 compared to their L1. Yet, the latter difference is extremely slight across languages and may be due to other factors generating variability in the total duration of the interaction.

This visualisation tool is particularly helpful for observing single interactions with a high degree of detail. However, larger amounts of data require a synthetic representation that is easier to interpret. Therefore, in addition to this figure, the Praat script derives a table used to generate pie plots (example in Figure 3.7) in R

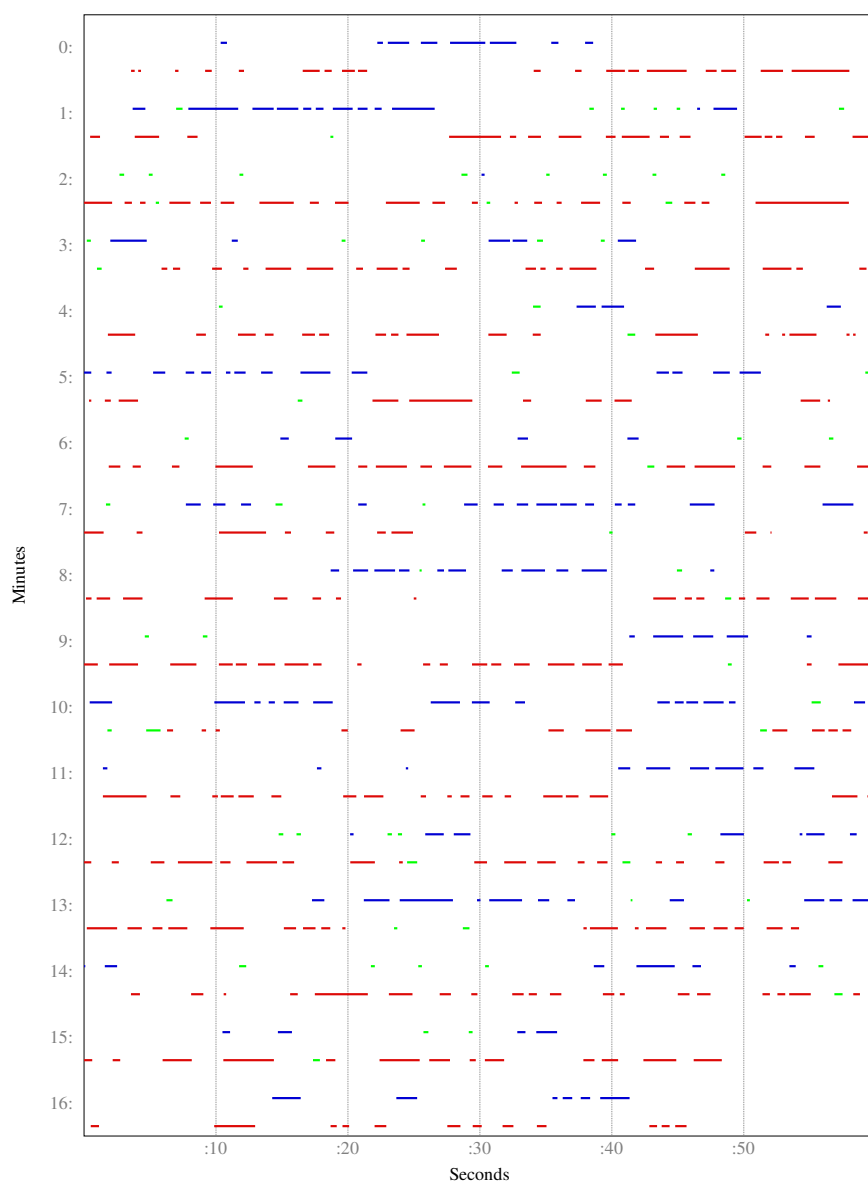


Figure 3.4: Visualisation of the interactional flow in L2 German (dyad with low L2 proficiency). Color code identifies speakers according to their role in the Map Task: red – instruction giver, blue – instruction follower. Green bars mark backchannels.

3 Turn-taking fluency in L2 interactions

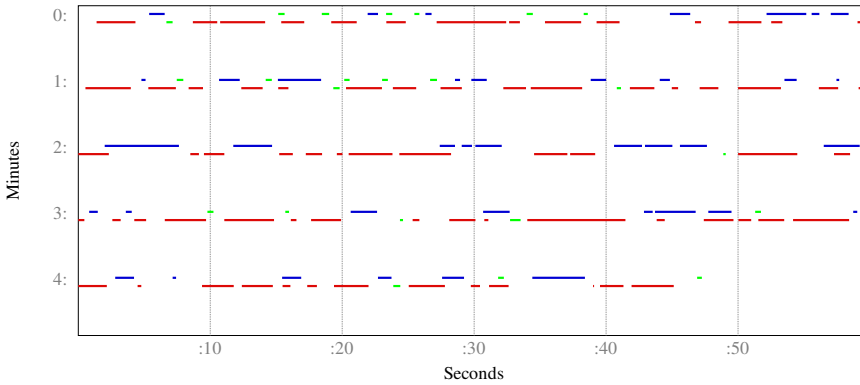


Figure 3.5: Visualisation of the interactional flow in L1 Italian (dyad with high L2 proficiency). Color code identifies speakers according to their role in the Map Task: red – instruction giver, blue – instruction follower. Green bars mark backchannels.

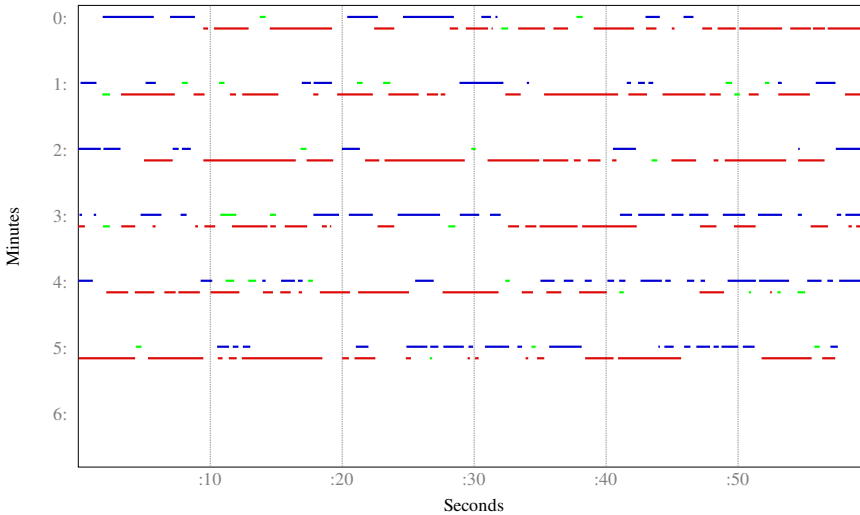


Figure 3.6: Visualisation of the interactional flow in L2 German (dyad with high L2 proficiency). Color code identifies speakers according to their role in the Map Task: red – instruction giver, blue – instruction follower. Green bars mark backchannels.

(R Core Team 2013) from the extracted data. The five sections of the pie plots use the same colour-coding as the conversation charts to show the percentages of speech uttered by each speaker. The radius of the circle represents the total duration of the interaction: the bigger the pie, the longer the interaction. While the conversation chart is useful for observing the time-aligned development of the interaction, this pie plot summarises and quantifies the partition of the interaction into the five classes of conversational activities and its total duration. Applied to L2 data, this plot is helpful for understanding to which extent high-proficiency L2 interactions resemble L1 conversational patterns more than low-proficiency L2 interactions in terms of changes in the proportions of conversational activities across L1 and L2. Finally, the extracted metrics can be used for hypothesis testing as discussed in the following paragraph.

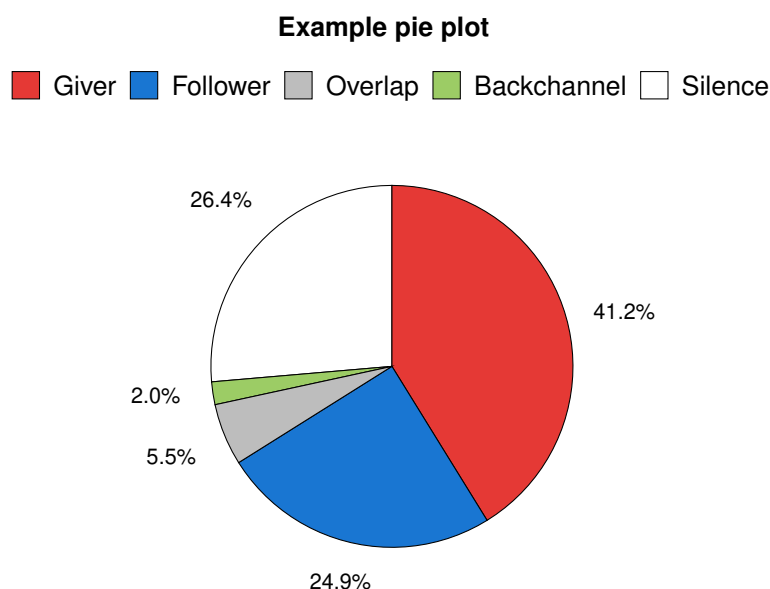


Figure 3.7: Example of pie plot summarising the five classes of conversational activities. The radius corresponds to dialogue duration.

3.2.5 Bayesian analysis

I statistically tested whether changes in the proportions of classes of conversational activities can predict learner proficiency. According to our expectation, with increasing proficiency, learners should approach their own native conversational patterns when speaking in their L2.

3 Turn-taking fluency in L2 interactions

Bayesian hierarchical linear models were fitted using the Stan modelling language (Carpenter et al. 2017) and the package *brms* (Bürkner 2016). For each language group, the differences in speech time, silence, overlap and dialogue duration were tested as a function of the factor PROFICIENCY (reference level is ITALIAN L1).

For the classes of conversational activities (i.e. speech time, silence, overlap), proportion values were taken into account. Therefore, a zero-inflated beta distribution was used and priors for the intercept and the regression coefficient were defined based on data exploration. For speech time of giver, the intercept was set at $\mu = 0$, $\delta = 0.3$ and the regression coefficient at $\mu = 0$, $\delta = 0.55$; for speech time of follower, the intercept was set at $\mu = 0$, $\delta = 0.3$ and the regression coefficient at $\mu = 0$, $\delta = 0.1$; for silence, the intercept was set at $\mu = 0$, $\delta = 0.4$ and the regression coefficient at $\mu = 0$, $\delta = 0.2$; for overlap, the intercept was set at $\mu = 0$, $\delta = 0.06$ and the regression coefficient at $\mu = 0$, $\delta = 0.035$. In all models, I used a beta distribution with $\alpha = 1$ and $\beta = 1$ for the alpha parameter (i.e. the probability of an observation being 0 or 1), a beta distribution with $\alpha = 1$ and $\beta = 1$ for the gamma parameter (i.e. if the probability an observation is 0 or 1, the probability being 1), a gamma distribution with $k = .01$, and $\theta = .01$ for the phi (precision) parameter. The default settings of the *brms* package were retained for all other parameters. For total duration of dialogue, the total amount of time samples extracted by Praat were considered. Therefore, a lognormal distribution was used and priors for the intercept and the regression coefficient were defined based on data exploration. The intercept was set at $\mu = 0$, $\delta = 5000$ and the regression coefficient at $\mu = 0$, $\delta = 10000$. The default settings of the *brms* package were retained for all other parameters. Three sampling chains for 4000 iterations with a warm-up period of 3000 iterations were run for all models. There was no indication of convergence issues (no divergent transitions after warm-up; all Rhat = 1.0).

The expected values under the posterior distribution and their 95% credible intervals (CIs) are reported for all relevant contrasts (δ), i.e. the range within which an effect is expected to fall with a probability of 95%. For the difference between each contrast, the posterior probability that a difference is bigger than zero ($\delta > 0$) is also reported to ensure comparability with conventional null-hypothesis significance testing. In particular, it is assumed that there is (compelling) evidence for a hypothesis that states $\delta > 0$ if zero is (by a reasonably clear margin) not included in the 95% CI of δ and the posterior $P(\delta > 0)$ is close to one (cf. Franke & Roettger 2019).

All models, results and posteriors can be inspected in the accompanying RMarkdown file in the OSF repository (<https://osf.io/9ca6m/>).

3.3 Quantitative analysis

Data visualisation (Section 3.2.4) using conversational charts has suggested that a higher proficiency level in L2 enables a degree of smoothness in managing the interactional flow that is closer to the one learners show in their native language, possibly due to an enhanced automatization of the cognitive processes required to speak a foreign language, while the pie plot summarises the changing of the interactional patterns with increasing command of the L2.

To test informativeness, four explorative pie plots representing dialogues performed by two dyads with different proficiency in L1 and L2 are displayed in Figure 3.8. On the left, there is a dyad with low L2 proficiency and on the right, a dyad with high L2 proficiency (same speakers as in the conversation charts, Figures 3.3–3.4 and Figures 3.5–3.6 respectively). The high-proficiency dyad (advanced – C1 level) presents two very similar patterns of interaction across languages. The ratio of time speaking between the giver and the follower remains approximately 3:1 when they repeat the task in the L2. In contrast, the low-proficiency dyad (beginner – A1 level) presents two very different interactional patterns. In the L1, the ratio of time speaking between the giver and the follower is 3:1, whereas in the L2 the ratio changes to 2:1, with the giver speaking less in the L2 than in the L1.⁶ Furthermore, more than half of the conversation consists of silence. Commonalities across proficiency levels seem to be relative to a longer total duration of the L2 dialogue, and a reduced amount of overlap in the L2, whereas in both dyads the amount of speech time of the follower remains unchanged across languages.

Statistical testing of these observations confirms that speech time of the giver as well as silence are good predictors of learner proficiency in terms of similarity to their native baseline. For both metrics, beginners are robustly different from their L1 Italian, in contrast to advanced learners. Specifically, beginner instruction givers speak notably less in their L2 than in their L1 ($\delta = -0.16$, CI [-0.29; -0.04], $P(\delta > 0) = 0.99$), and silence in their interactions is remarkably more prevalent than in the L1 ($\delta = 0.23$, CI [0.08; 0.39], $P(\delta > 0) = 1$). In contrast, within the advanced group, these two parameters show no robust differences between L1 and L2 (for giver's speech time: $\delta = -0.1$, CI [-0.23; 0.04], $P(\delta > 0) = 0.88$; for silence: $\delta = 0.13$, CI [-0.02; 0.28], $P(\delta > 0) = 0.93$). As observed in data exploration, speech time of the follower did not show robust variation across L1 and L2, for any proficiency level. The same holds true for overlap, in contrast with preliminary observations. Finally, the total duration of the dialogue was found to be

⁶An overall dominance of the instruction giver is expected due to the nature of the task itself.

Exploratory data

Proportions of conversational activities across proficiency levels

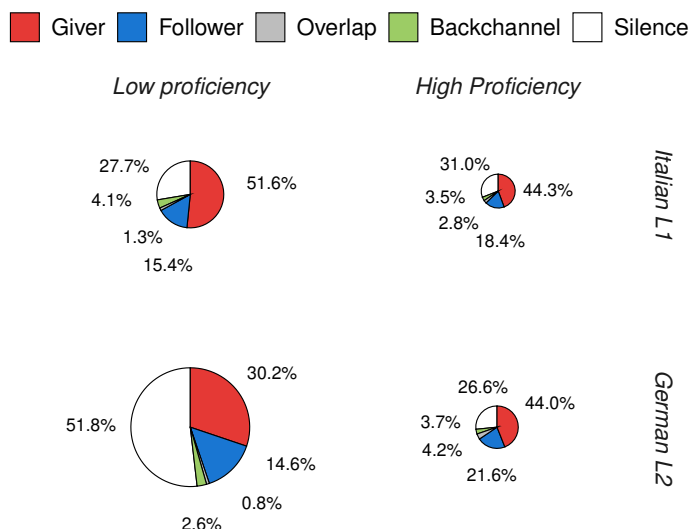


Figure 3.8: Pie plots summarising conversational activities for a low-proficiency dyad (IF, on the left) and a high-proficiency dyad (BS, on the right). Plots in the top row display dialogues in the L1 and plots below dialogues in the L2. The radius of each pie corresponds to dialogue duration.

robustly longer in L2 than in L1, independently of learner proficiency (for beginner learners: $\delta = 1894.26$, CI [920.15; 2979.22], $P(\delta > 0) = 1$; for advanced learners: $\delta = 1381.85$, CI [512.03; 2436.16], $P(\delta > 0) = 1$). Therefore, the three latter conversational metrics did not turn out to be good indicators of the Italian learners' L2 proficiency levels based on their oral performance.

Data averaged across dyads (Figure 3.9) visually support these results in showing that silence gradually decreases and speech time of the giver gradually increases across the two proficiency levels as compared to the L1. To examine the whole corpus, Figure 3.10 depicts the pie plots for all learners' interactions displayed by increasing proficiency in both their L1 and L2.⁷ It can be noticed that:

- L2 interactional patterns in the beginner group are highly variable across L1 and L2 (with the exception of dyads CV, AR and CC);

⁷All by-dyad pie plots are presented singularly, with the percentages relative to the conversational activities, in the Appendix, for the sake of improved visualisation. L1 Italian and L2 German dyads are displayed in Figures A4–A23, while L1 German in Figures A24–A32.

- L2 interactional patterns in the advanced group start consistently resembling those produced in L1, especially from dyad RS on;
- Amount of silence is consistently higher in beginner L2 interaction compared to their L1, whereas in the advanced group some dyads present very similar silence values across L1 and L2 (see CE, RC, BS and FF).

Averaged proportions of conversational activities across groups

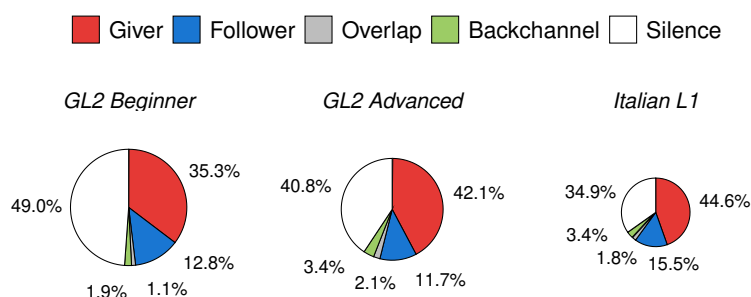


Figure 3.9: Averaged pie plots summarising conversational activities for beginner and advanced learners, and their native Italian baseline. Radii correspond to average dialogue duration.

3.4 Qualitative analysis

For a statistical analysis, larger and homogeneous groups of learners were necessary, thus a regrouping into two main proficiency groups was performed. Nevertheless, changes in interactional patterns might be already visible at the more fine-grained CEFR levels. Thus, after having demonstrated the application of the method to L2 data, I will now deepen the discussion of the results obtained, going beyond a by-group analysis and providing a qualitative analysis of learners' improvement across CEFR levels, with a special focus on dyad-specific behaviour. Additionally, based on the CEFR levels, I will investigate whether the results related to interactional competence align with those from the lexical competence test, specifically examining whether the development of various abilities within overall communicative competence occurs in parallel or not.

3.4.1 Procedure

To investigate changes in interactional patterns at the more fine-grained CEFR levels, I used a measure that summarises learners' interactions across their L1 and

3 Turn-taking fluency in L2 interactions

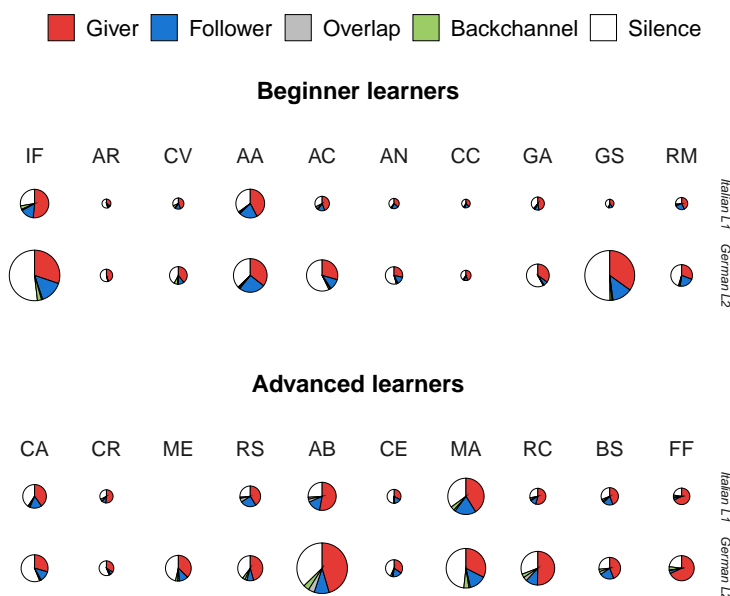


Figure 3.10: Pie plots summarising conversational activities for all beginner and advanced learners in both their L1 Italian and L2 German. Plots in the upper row display dialogues in L2 and plots in the bottom row dialogues in L1. Pairs of letters identify dyads. The radius of each pie corresponds to dialogue duration.

L2 for each dyad, i.e. the difference between L1 and L2 ratios of time speaking of the giver on the follower. The reason is that interaction performance should not be analysed or evaluated in isolation, i.e. by speaker, as it is a form of co-creation of the discourse by both interlocutors.

Before performing this calculation, the time proportion of backchannels was reassigned to the relative speaker's speech time in order to obtain the absolute speech time for each speaker. Then, the difference between L1 and L2 ratios of speaking time between the giver and the follower was calculated, so that each data point presented in the results represents a dyad and not a single speaker. During data exploration, it was noticed that the difference between the L1 and L2 ratios of speech time between the giver and the follower is around 0 for high-proficiency learners and different from 0 for less proficient learners. Therefore, as a result of the difference between these two ratios, points approaching 0 will indicate less difference in the interactional behaviour of learners across languages.

3.4.2 Results

Results are presented in two versions: the first one including all data points (Figure 3.11) and the second one excluding an outlier for the sake of improved visualisation (the golden data point in Figure 3.11 is not present in Figure 3.12).

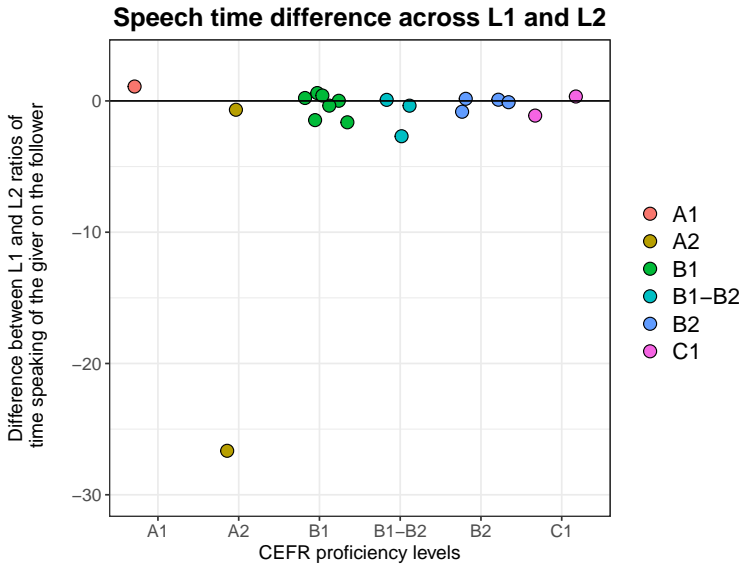


Figure 3.11: Difference between L1 and L2 ratios of time speaking of the giver on the follower. The graph includes all data points. The x-axis displays the proficiency level, the y-axis shows the values resulting from the formula. The more points approach 0, the less difference there is between learners' interactional behaviour in L1 and L2.

The graph reveals minor variation among dyads. Two out of the three dyads with A1 and A2 proficiency (considered as “beginner” levels in the CEFR) are not as far away from 0 as some B1 and B1-B2 dyads (considered as “intermediate” levels in the CEFR), showing that the difference in learners' interactional behaviour in beginners can be less than in intermediate learners. Indeed, while proficiency assessment is mostly based on testing grammatical and lexical resources, these are not the only factors contributing to conversational rhythm. Many other linguistic and extralinguistic factors can play a role in goal-oriented conversation, such as personality, engagement in the task, relationship between speakers, and not to forget the skill to strategically draw on the few resources beginners have to reach the goal of the interaction. These two A-level dyads might be an example of the latter case. However, with few samples for the A-level of proficiency, this result cannot be considered informative about a general trend.

3 Turn-taking fluency in L2 interactions

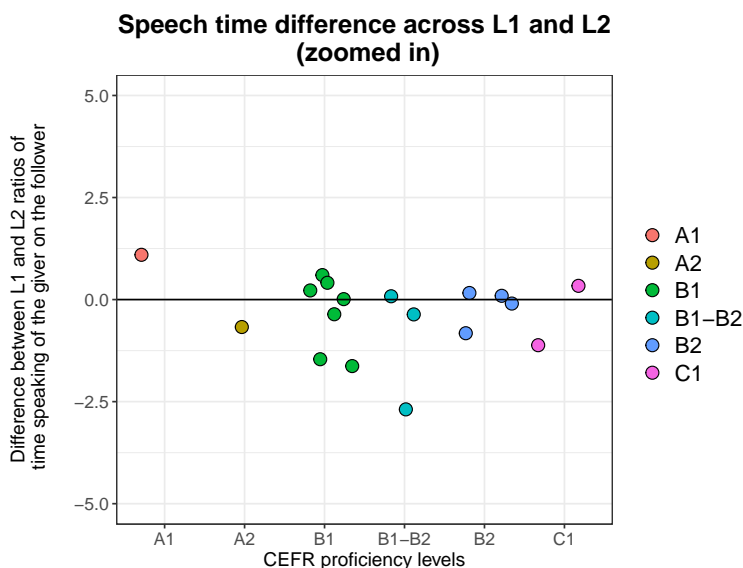


Figure 3.12: Difference between L1 and L2 ratios of time speaking of the giver on the follower. The graph does not include one outlier for the sake of a better visualisation. The x-axis displays the proficiency level, the y-axis shows the values resulting from the formula. The more points approach 0, the less difference there is between learners' interactional behaviour in L1 and L2.

The third A-level dyad, i.e. the golden outlier, shows a particular behaviour: the follower only utters a few sentences towards the end of the dialogue in both the L1 and L2. In such cases, using the L1 as a baseline for learners' interactional behaviour is particularly beneficial. The fact that the follower does not contribute to the conversation in L2 German would have generally been associated with poor command of the L2, but this speaker behaves exactly the same in L1.

On the other hand, the variability displayed in B1 and mixed B1-B2 levels of proficiency tends to be reduced in the more advanced B2 and C1 levels (C-levels are considered as "advanced" levels in the CEFR), which show a narrower range of values. This observation is in line with the statistical analysis, revealing that learners with a higher proficiency, in particular from a B2 level, produce a highly similar interactional pattern in their L1 and L2. In other words, less cognitive load, time for information retrieval and formulation, and attention required in the L2 due to a higher degree of automatization allow learners to achieve an increased degree of smoothness in terms of interactional flow. Still, for the C1 level, too, there is a limited number of dyads. Therefore, a more conspicuous and

homogeneously distributed number of samples across proficiency groups would be required to confidently test the trend observed across the CEFR proficiency levels.

As mentioned in Section 1.3.2, learners also took part in an online test for lexical competence, i.e. the German version of LexTALE (Lemhöfer & Broersma 2012). Figure 3.13 explores the relation between lexical competence and the interactional pattern. The range of values on the y-axis is reduced for the sake of a clearer visualisation, cutting out the golden outlier as in the previous figure. The radius of the circles represents the score obtained in the test for vocabulary knowledge by the giver, who is the one leading the task and contributing the most to the conversation. The larger the circle, the higher the lexical score the giver received. Lemhöfer & Broersma (2012: 341) report a correspondence of A1, A2 and B1 levels to scores below 59, B2 level to scores between 60 and 80, and C1 and C2 levels to scores between 80 and 100. Based on their data, the lexical competence of these learners does not seem to increase in parallel with their overall proficiency level as we have no score above 65. Regarding the relation of the lexical competence score to interactional patterns, the graph does not show a clear trend. In the B1, B1-B2 and B2 groups, both very low and high lexical scores are near to zero, suggesting that lexical competence does not seem to be a factor determining how different the interactional patterns are in L1 and L2. This provides some evidence for the previous statement that in a goal-oriented cooperation task, skills other than merely linguistic ones come into play, and a good strategical competence can compensate for the low level of L2 vocabulary and grammar knowledge.

3.5 Cross-linguistic comparison

After examining the turn-taking behaviour of Italian learners, an overview of L1 German conversational management patterns and their comparison to L1 Italian is appropriate. Exploring potential cross-linguistic differences may reveal critical aspects for learners that should be addressed in SLA.

Averaged values for the five classes of conversational activities (Figure 3.14) show that German L1 dialogues have a much longer duration than those performed by Italian native speakers. Indeed, one common strategy among L1 German dyads was to check the drawn path at the end of the task once again, which shows careful consideration regarding the correct execution of the task. Furthermore, silence is slightly reduced in favour of speech time of the follower, which suggests a more interactive and collaborative strategy than in Italian L1 conversations, despite the predetermined roles.

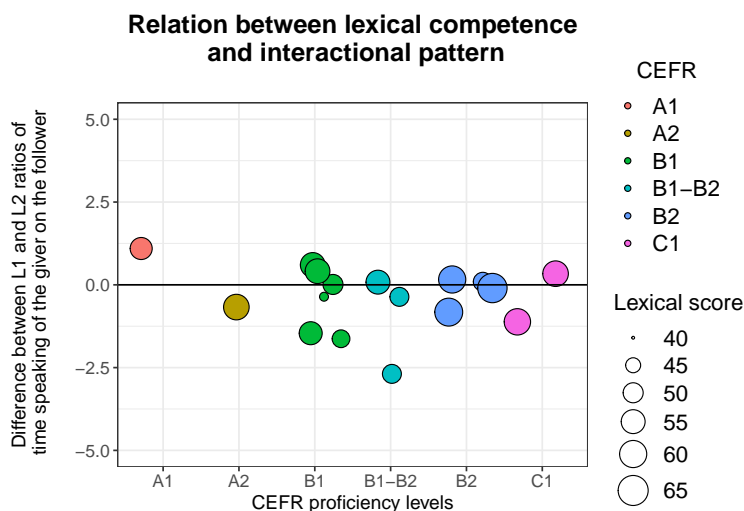


Figure 3.13: Relation between lexical competence and interactional pattern. The x-axis displays the proficiency level, the y-axis shows the difference between L1 and L2 ratios of time speaking of the giver on the follower. The radius of the circles represents the score for L2 lexical competence of instruction givers.

A closer look at by-dyad data (Figure 3.15)⁸ confirms the large difference in total dialogue duration. Only one German L1 dyad, SI, presents a dialogue duration which falls within the range of Italian L1 duration values. Moreover, there seems to be more consistency across German dyads in the proportions of classes of conversational activities than in Italian dyads, even if this result might be due to the smaller sample. For this same reason, these observations have to be taken as preliminary. They suggest some cross-linguistic differences which might be interpreted as being culture-specific interactional conventions and are worth being investigated further.

3.6 Conclusion

In this study, I have problematised the absence of a standardised instrument for the quantification of interactional competence in L2. To open up new perspectives for L2 assessment, I presented visualisation tools and a quantification method that can extract reliable and testable metrics for interactional aspects of

⁸All by-dyad plots are presented individually in the Appendix for improved visualisation of percentage values (Figures A4–A32).

Proportions of conversational activities in L1 by group

Giver Follower Overlap Backchannel Silence

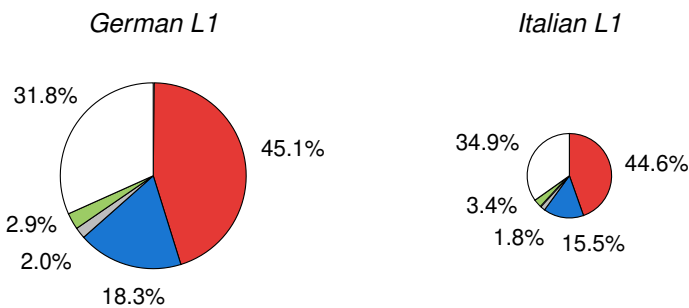


Figure 3.14: Averaged pie plots summarising conversational activities for native German and Italian speakers. Radii correspond to average dialogue duration.

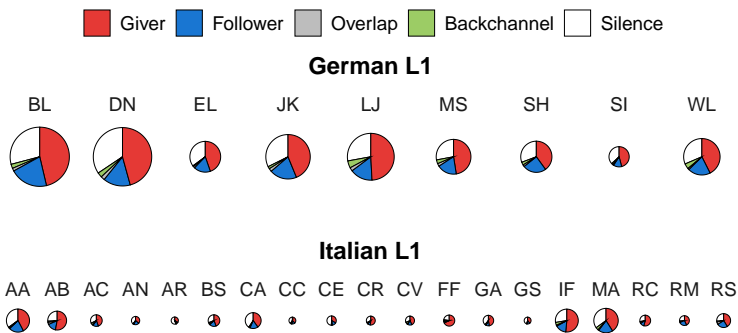


Figure 3.15: Pie plots summarising conversational activities for all L1 Italian and L1 German dyads. Plots in the upper line display dialogues in L1 German and plots in the bottom line dialogues in L1 Italian. The radius of each pie corresponds to dialogue duration.

3 Turn-taking fluency in L2 interactions

communication, i.e. speaking time of participants, silence, overlap, backchannels and total duration of the dialogue. The informativeness of this method with regard to learners' L2 proficiency was tested on a corpus of L1 and L2 interactions by conducting data exploration and subsequent hypothesis testing.

Results show that more proficient learners better maintain the natural interactional rhythm they have in their L1 in the L2, which is especially clear in the speech time metrics for instruction giver and the total amount of silence. A further qualitative analysis suggests that improvements might be already visible at the more fine-grained CEFR levels, but this observation requires further testing on a larger scale considering that this corpus only includes a few samples for the A- and C-levels of the proficiency scale, in contrast to the more conspicuous B-level group.

Lexical competence did not seem to influence learners' interactional behaviour, which suggests that mastering the lexicon does not automatically ensure a higher degree of success in oral interactions. Moreover, lexical scores did not seem to lead to the corresponding levels of general L2 competence. Indeed, the learning process is not linear, and there is no discrete order for L2 knowledge acquisition (Nava 2010), so that different skills can improve at different speeds. Since open interactional tasks test learners' L2 abilities in a more comprehensive way, an enhancement of these kind of tasks in L2 experimental and testing settings can help obtain a clearer picture of learners' L2 general proficiency, and possibly shed light on the interplay among different skills.

In addition, a preliminary comparison of L1 German and L1 Italian conversations suggests some likely cross-linguistic/-cultural differences in the total duration of dialogues as well as the speech time of the instruction follower, which seem to suggest a more careful and collaborative approach by L1 German compared to L1 Italian dyads. In order to relate these results to language pedagogy, the differences in interactional conventions across languages/cultures deserve being investigated further on a larger scale.

The method proposed has considerable potential for the analysis of L2 oral interactions. It permits an immediate comparison of learners' interactional behaviour in the L1 with their performance in the L2, both in a detailed and synthetic way throughout the different stages of the learning process, which can be complemented by statistical testing. Moreover, being performed mostly in an automated way, this is a labour-saving analysis. For these reasons, this method of visualisation and quantification of oral interactions could represent a starting point for quantifying L2 interactional competence based on interactional fluency in a standardised way.

At the same time, there are some limitations that need to be considered. First, this analysis offers only a partial, structural view of the interaction as it is content-free and based on temporal measures only. The missing verbal content is crucial for understanding the underlying reason generating the metrics. As an example, the verbal content is necessary for clarifying the different nature of silence and discourse chunking across the L1 and the L2. Consider silence following hesitant feedback by the interlocutor signaling a lack of understanding. The primary speaker might want to give time to the interlocutor to explain the nature of their hesitation with the unclear content, or take time to reformulate their own message in a different way. Only in the latter case can the amount of silence be interpreted as a measure of cognitive load and related to a less complete mastery of the language. Nevertheless, the implementation of other content-related information is possible by integrating the conversational activities with categories taken from the pragmatic and strategic skills indicated in the CEFR interaction scale (e.g. asking for clarification, compensating, cooperating, monitoring and repair), which would add more layers of information and complexity. Another option would be to integrate content-related information at turn transitions, i.e. turn sequences in which speakers alternate in occupying the floor, to assess the strategies learners use to coordinate the interaction throughout the learning process.

Secondly, there are non-verbal forms of communication (e.g. eye gaze, gesture, posture) which also contribute to the interactional patterns (see, for example, Kosmala 2024), but were not captured by the present data collection method optimised for verbal communication. Thus, the results should be interpreted within the context of the data collection setting. A future integration of non-verbal communicative features into this tool may be possible if learners consent to being video-recorded. The time investment necessary for such an analysis should be also tested in order to evaluate the practicality of its application.

Lastly, I have proposed a by-dyad quantification method which, together with the suggestion of collecting data through spontaneous peer conversations, contrasts with the need to assess each learner's L2 competence individually. In real-world scenarios, such as language certification settings, learners' L2 competence is typically assessed on an individual basis to provide a final score. However, in interactions, both interlocutors share responsibility for co-constructing the communicative exchange. Therefore, further development of this tool, which was designed for dyadic exchanges, should account for each speaker's contribution to the interaction.

Overall, this study shows that it is possible to reliably identify conversational activities related to language proficiency, opening up possibilities for future im-

3 Turn-taking fluency in L2 interactions

plementations based on the linguistic resources contributing to these activities. One linguistic feature that plays a significant role in interactional fluency and might enrich these implementations with content-based data is the use of feedback as a measure of active listening, which will be the focus of the next chapter.

4 Backchannels in L2 interactions

This chapter focuses on examining a particular ability of L2 interactional fluency, specifically the use of vocal feedback signals (henceforth “backchannels”) in two-party conversations.¹

Backchannels are very short lexical or non-lexical utterances used by a listener to signal acknowledgment to what the speaker is saying. Given their role, they support the ongoing turn of the interlocutor, and positively contribute to a smooth turn alternation and, eventually, structure in a dyadic conversation. While these small tokens are generally unnoticed in a conversation among native speakers, they can stand out in multicultural and multilingual settings if aspects of their realisation do not follow the native norm and potentially cause misunderstanding.

Given the lack of a comprehensive study on backchannel production across languages and in L2, I propose a study with a within-subjects design in order to investigate several aspects of backchannel use across L1 and L2 and the possible relation between them. The goal is to highlight critical aspects of backchannel production in SLA.

4.1 Background

4.1.1 Backchannel contribution to interactional fluency

As discussed in the previous chapter, one issue in SLA research has been the question of how to assess communicative competence in a quantitative and systematic way, while taking into account idiosyncratic and contextual factors impacting the L2 learning process and L2 oral performance. Fluency has been widely recognised as a central aspect in the assessment of L2 oral proficiency (De Jong 2016), but most studies have focussed on individual measures of L2 fluency, while the majority of real-life oral performances are interactions and much less often monologues.

¹The analysis provided in this chapter integrates work previously published in Sbranna et al. (2022, 2023, 2024). Its inclusion in this book enables a discussion of the findings within the broader framework of interactional fluency and their potential pedagogical applications.

Fluency in dialogue is highly determined by the specific interaction mechanism that arises between the two parties of a conversation, so that along with individual factors, unique dyad-related factors play a fundamental role (for an extensive discussion on this topic see also Sbranna et al. 2020). For this reason, interaction has been described as a co-construction process both in research (Hall 1995; He & Young 1998; Jacoby & Ochs 1995; McCarthy 2009) and in the CEFR (Piccardo et al. 2018). The smoothness of a conversation is achieved, among other factors, through the rhythm of turn-taking (Sacks et al. 1974). Indeed, smooth or disfluent transitions of turns can take place at turn-boundaries, and interlocutors have to appropriately foresee the end of the other party's turn and react quickly and accordingly (Bögels & Torreira 2015; Levinson 2016). Despite usually going unnoticed in conversation (Shelley & Gonzalez 2013), one important linguistic means that can facilitate turn transitions is the use of so-called “backchannels”.

Backchannels are very short lexical and non-lexical utterances, like *okay* or *mm-hm*, which have traditionally been described as non-intrusive tokens – that is, as not claiming a floor transfer – used to signal the listeners' active engagement, showing acknowledgement and understanding (Schegloff 1982; Yngve 1970: 19). By supporting the ongoing turn of the interlocutor, backchannels positively contribute to fluency in social interactions (Amador-Moreno et al. 2013) as they maintain flow and contribute to creating structure in a dyadic conversation (Kraut et al. 1982; Sacks et al. 1974; Schegloff 1982).

On the other hand, backchannels can be potentially misleading in cross-cultural contexts where different culturally-shaped communicative conventions come into contact (Cutrone 2005, 2014; Ha et al. 2016; Li 2006). Research has indeed provided evidence for language- or variety-specific backchannel characteristics concerning duration, frequency, location, intonation and function, and these are a possible source of negative social implications in a communicational setting in which the interlocutors' linguistic backgrounds diverge.

For these reasons, backchannels in L2 learning are extremely important. The CEFR (Figueras et al. 2009) lists the use of feedback expressions under passive competence already at the A2 level. However, backchannels are not explicitly thematised in most L2 classrooms, and it cannot be taken for granted that learners acquire appropriate backchannelling behaviour solely through exposure to the target language. Moreover, teachers are not always native speakers and input on this particular interactional feature might be completely absent from classroom settings.

Against this background, two possible manifestations of backchannels in interlanguage can be expected. On the one hand, it is possible to assume that backchannels go unnoticed in conversation, resulting in a transfer of features

from the L1 to the L2. On the other hand, assuming that there is exposure, backchannels might be perceived by learners as salient features of foreign speech and receive an appropriate level of attention, which would favour an adaptation to target language patterns. In the latter case, a more target-like backchannel behaviour should be observed, especially at an advanced level, i.e. with more experience of and exposure to the target language. With these two scenarios in mind, I will explore the use of backchannels in L2 learning.

4.1.2 Backchannel definitions and categorisations

In the literature, there is little agreement about the definition of backchannels (as noticed by Lennon 1990, 2000; Rühlemann 2007; Wolf 2008 among others).

In his analysis of telephone conversations, Fries (1952) was probably the first to recognise these “signals of attention” that do not interrupt the speaker’s talk. Since then, other terms have been used to define this phenomenon, such as “accompaniment signals” (Kendon 1967), “receipt tokens” (Heritage 1984), “minimal responses” (Fellegly 1995), “reactive tokens” (Clancy et al. 1996), “response tokens” (Gardner 2001), “engaged listenership” (Lambertz 2011) and “active listening responses” (Simon 2018).

The term “backchannel communication” was first coined by Yngve (1970) to define the channel of communication used by the listener/recipient to give useful information to their interlocutor without claiming a turn, in opposition to the main channel used by the speaker holding the floor.

Initial investigations into backchannelling primarily focused on American English (Duncan 1974; Duncan & Fiske 1977; Fries 1952; Goodwin 1986; Jefferson 1983; Schegloff 1982; Yngve 1970). These pioneering works sought to establish a definition of backchannels and proposed classifications of backchannel types grounded in either their pragmatic function or formal realisation.

Schegloff (1982) noted that these short utterances were mainly used by the listener not only to acknowledge the interlocutor’s turn, but also to invite the primary speaker to carry on with his turn. For this reason, he defined the minimal utterances used in the specific contexts of an ongoing turn by the interlocutor as “continuers”. Jefferson (1983) introduced the term “acknowledgement tokens”. Indeed, the term backchannel in its narrow use refers to tokens used to signal acknowledgement and understanding of what the interlocutor is saying, while inviting the main speaker to continue (Beňuš et al. 2007; Hasegawa 2014).

In its broader use, the term “backchannel” has also been matched to numerous other functions, and some attempts at establishing a function-based categorisation have been made. For example, Jefferson (1983), Drummond & Hopper (1993)

and later Jurafsky et al. (1998), Savino (2010), Savino (2011), Savino (2014), Savino & Refice (2013) further distinguish acknowledgement tokens marking “passive reciprocity”, as in the case of continuers, from those marking “incipient speakership”, signalling a listener’s intention to start a turn of their own. Senk (1997) categorises backchannels according to the functions of continuer, understanding, agreement, support, strong emotional answer and minor additions. Kjellmer (2009) recognises five functions of backchannels: regulative, supportive, confirmatory, attention-showing and empathetic. Tolins & Fox Tree (2014) distinguish context-generic backchannels, used as continuers and promoting the production of new information, and context-specific backchannels, also called “assessments” in previous studies (Goodwin 1986), such as *really* or *wow*, eliciting further elaboration of what has just been said.

As far as their formal realisation is concerned, backchannels present a high degree of lexical variability, although they can also be non-lexical, e.g. realised through vocal noises (Wong & Peters 2007), and non-verbal, making use of visual modalities such as facial expressions, head movements, and gestures (Tolins & Fox Tree 2014), and responsive laughter (Hasegawa 2014). Some structurally motivated proposals of classifications have been advanced to categorise backchannel lexical realisations. Tottie (1991) classifies them into simple, double, and complex types. Simple backchannels are composed of one single utterance, e.g. *yes*, double backchannels are repeated simple types, e.g. *okay okay*, and complex backchannels are a combination of different simple types, such as *okay yes right*. Wong & Peters (2007) differentiate between minimal, lexical, and grammatical types. Minimal types are defined as non-lexical items that are semantically empty and items expressing polarity, e.g. *mmhm*, *yes*, and *no*. Lexical types are considered to be all single words that are codified in dictionaries and show an increase in semantic weight, such as *really*, *right*, and *good*. Finally, by grammatical types they mean predication in the form of short codified phrases, such as *I see*, brief questions, repetitions, sentence completions, and commentaries.

As noticed by Edlund et al. (2010), the variety of names and categorisations provided by previous studies are often vague and overlapping. Moreover, the labelling schemes adopted treat backchannels quite differently. Faced with these difficulties, the authors propose a more general unit called “very short utterance” (VSU) to capture the large range of interactional dialogue phenomena commonly referred to as backchannels, feedback, and continuers.

Subsequent research broadened its scope beyond American English, examining other languages and revealing variations in backchannel usage across cultures and languages (Berry 1994; Clancy et al. 1996; Cutrone 2005, 2014; Heinz 2003; Kraaz & Bernaisch 2022; Li 2006; Nurjaleka 2019; Tao & Thompson 1991;

Tottie 1991; Ward & Tsukahara 2000). These cross-cultural and cross-linguistic differences will be further explored in the following section.

4.1.3 Backchannel use across languages and cultures

One focus in the field of backchannel research has been variation across languages and cultures. Differences in backchannel use have been identified regarding their frequency, duration, location, lexical types, functions, and intonation.

Because it is influenced by cultural norms, backchannelling has been found to vary even among varieties of the same language. Tottie (1991) reports differences with regard to frequency and types across American and British English, showing that in American conversations there was an average of sixteen backchannels per minute, compared with just five backchannels per minute in British conversations. Similarly, differences were observed across Sri Lankan and Indian English in type, frequency, and function (Kraaz & Bernaisch 2022).

Some studies report the impact of different backchannelling behaviour on the turn-taking system. For instance, in a cross-linguistic study on Spanish and North-American English, Berry (1994) found that backchannels were more frequent and longer among Spanish speakers, resulting also in longer stretches of overlapping speech. In turn, American English speakers were shown to use more overlapping backchannels than Germans, as reported in a comparative study by Heinz (2003).

The differences found lead to the hypothesis of a potentially negative effect on communication in intercultural conversations. In a study on responsive tokens in English, Mandarin and Japanese, Clancy et al. (1996) observed that Japanese speakers produced the most frequent reacting tokens, placing them in the middle of the interlocutor's speech. Mandarin speakers, in contrast, produced the fewest backchannels, and mostly at TRPs, i.e. at the end of the interlocutor's turns. American English was in the middle between the two language groups in terms of frequency, and reacting tokens were placed both within an interlocutor's turn and at TRPs, but preferably at grammatical competition points. The authors speculate that, in Japanese, backchannels are used as a form of emotional support and cooperation, whereas, on the opposite pole, Mandarin speakers might perceive Japanese backchannels as intrusive in comparison to their tendency not to interrupt the other speaker out of respect. American English speakers, likewise, might find Japanese speakers disruptive. However, the scarce reactions of Mandarin speakers would leave them wondering what their listeners are thinking (Clancy et al. 1996: 383). Similar hypotheses were tested in a study on backchannel intonation, in which Ha et al. (2016) found differences

across Vietnamese and German. While Vietnamese continuers are consistently level or falling, German equivalents are tendentially rising. Based on the results of a previous perception experiment (Ha 2012), the authors hypothesise probable misunderstandings in intercultural dialogues. In Vietnamese, rising pitch as used by Germans might be interpreted as impolite. Conversely, for German natives, the level/falling pitch used by Vietnamese might cause irritation (Stocksmeier et al. 2007) and could be interpreted as showing disinterest, or as an attempt to end the interlocutor's turn.

Given the observed differences across languages, the immediate next step in research was to put the consequences of this variation in intercultural conversations to the test and find out whether and to what extent differences in backchannel use can lead to miscommunication and/or have negative social implications. Li (2006) conducted a study on Canadian and Chinese speakers in intra- and intercultural conversations and showed that backchannels facilitate content communication among speakers of the same language. But when Canadian speakers were paired with Chinese speakers, the opposite effect was observed, leading to the claim that backchannel responses can be misleading in intercultural conversations and cause miscommunication. It was also found that Chinese speakers produced the most backchannels and Canadians the fewest, but when crossed, speakers tended to produce a number of backchannels in between. In a follow-up study providing an analysis of backchannel types (Li et al. 2010), it was found that, in intercultural conversations, both Canadian and Chinese speakers used other backchannels than in their respective native languages, showing some degree of speech convergence for both frequency and lexical type.

However, accommodation in intercultural conversations does not always take place automatically, with knowledge of language- and culture-specific conventions probably being essential. For example, a study reported that Japanese speakers did not adapt their active listening style in conversations with Americans, while Americans did, because “they clearly have the linguistic ability to do so” (White 1989: 74), suggesting that language proficiency might play a role for accommodation to take place. A high level of L2 proficiency can, indeed, provide the speaker with diverse linguistic means which can be selected according to context and the flexibility to recognise and switch among linguistic conventions.

4.1.4 Backchannels in L2 speech

To date, only relatively few studies have investigated backchannels produced by L2 learners. Their findings reinforce the assumptions made on the basis of

intercultural studies, showing that L1 backchannel behaviour is generally carried over to the L2, which can cause miscommunication and misperception.

For example, Cutrone (2005) examined the use of backchannels in dyadic interactions between Japanese EFL (English as Foreign Language) and British speakers. Differences were found in frequency, type and location, and affected intercultural communication negatively. The frequent backchannels used by the Japanese participants were interpreted as interruptions by the British speakers, and their interlocutors were perceived as impatient. In a follow-up study, Cutrone (2014) reports that Japanese EFL speakers used a greater number of backchannels because it helped them to feel comfortable as listeners, showing a behaviour similar to the one reported for L1 Japanese by Clancy et al. (1996).

Wehrle & Grice (2019) also report on the negative effect of transfer on intercultural communication. In a pilot experiment, they compared the intonation of backchannels in L2 German spoken by Vietnamese and observed that Vietnamese learners produced twice as many non-lexical backchannels (e.g. *mmhm*) with a flat intonation contour than German native speakers, showing a transfer from their L1. As noted earlier, in German, a flat intonation for backchannels can be perceived as disinterest and may cause offense (Ha et al. 2016).

Another study that hypothesises a transfer of backchannel features from the L1 to the L2 was conducted by Castello & Gesuato (2019). They investigated the frequency and lexical types of “expressions of convergence” (used synonymously with backchannels) in Chinese, Indian and Italian learners of English in a language examination setting. They found that Chinese learners used the most backchannels, and Indian learners used the least, while Italian learners showed a backchannel frequency in between these two groups. They also observed differences in the choice of backchannel types across groups, which was motivated by the influence of their own native language and culture.

A similar conclusion is reached by Shelley & Gonzalez (2013), who analysed backchannel functions in informal interviews in four English as Second Language (ESL) speakers with different L1 backgrounds and one American native speaker of English. They identify four backchannel functions: continuers (the listener is paying attention and gives the floor back), acknowledgements (the listener agrees or understands), newsmakers (the listener communicates an emotional reaction) and change of activity (the listener signals to move toward a new topic). They report an effect of culture-specific preferences as they find differences in the backchannel functions used across the four speakers: Japanese and Saudi Arabian speakers were found to use continuers, acknowledgements, and change of activity tokens; the Taiwanese speaker limited their use to acknowledgements; and the Egyptian speaker used both continuers and acknowledge-

ments; while the control American native speaker made use of the widest range of functions with continuers, acknowledgements, newsmakers, and change of activity tokens. These results are, however, difficult to interpret, as only one speaker is taken as representative of their language and culture, making it challenging to distinguish between language-specific and idiosyncratic factors.

Finally, there are studies showing that higher proficiency in the L2 implies a better ability to use backchannels. Galaczi (2014) compared the frequency of backchannels and expressions of confirmation among learners of English with different proficiency levels. The results showed that intermediate learners provided less feedback than highly proficient learners, among which the “ability to act as supportive listeners through backchanneling and confirmations of comprehension was found to be more fully developed” (Galaczi 2014: 570).

To summarise, previous research on various languages provides similar evidence 1) for miscomprehension and misperception of the interlocutor’s intentions due to a divergent use of backchannels from the native conventions, 2) for a transfer of native backchanneling behaviour to the L2, and 3) for proficiency as a positive factor for the improvement of learners’ L2 backchannelling ability.

At the same time, these studies have some limitations. Their results are not easily comparable as they differ considerably in design and methodology: how participants in the dialogue are matched, their status, their proficiency level in the language of the conversation, the setting of the dialogue, the method used for dialogue elicitation and aspects of backchannels analysed. Moreover, most studies have focussed on subjects with different L1 backgrounds, which is useful for detecting cultural-specific differences among groups of learners, but does not permit differentiation between transfer phenomena and cross-linguistic, speaker-specific characteristics.

Nevertheless, these findings have significant implications for the relevance of backchannels in language teaching environments. In order to better understand the mechanism behind cross-cultural backchannelling behaviour, it is important to shed light on how the backchannelling ability develops in interlanguages, with the goals of raising awareness in multicultural communicative contexts and improving L2 speakers’ interactional skills.

The aim of this study is to overcome some of the limitations mentioned by carrying out an in-depth analysis across languages and in an L2 with a homogeneous methodology. Using a within-subjects design, I investigate backchannel use across Italian learners’ L1 and L2 German and compare learners’ realisation of backchannels to a German native group to assess transfer phenomena and/or the acquisition of target-like backchannel features. In contrast to most previous studies, this investigation considers a broader range of backchannel features in

an effort to provide a more comprehensive picture of the phenomenon, including frequency, length, formal structure, (non-)lexical type, pragmatic function and intonation. Particular attention will be paid to dyad-specific behaviour in order to differentiate idiosyncratic factors from actual transfer or acquisition of patterns.

4.2 Method

For the purpose of this study, I will adopt the term “very short utterances” (Edlund et al. 2010) as a loose definition for the wide variety of interactional dialogue phenomena providing feedback to the interlocutors. I define backchannels as a specific class of VSUs with an acknowledging function, that is, showing understanding and acceptance of the interlocutor’s turn. This investigation also includes the VSU class of positive replies realised with the same token types as backchannels, such as *yes*, with the aim of assessing the impact of a different function on the distribution of lexical type and contour realisation. The criterion used to distinguish backchannels and positive replies is that backchannels are unsolicited, whereas positive replies are solicited by a yes-no or tag-question formulated by the primary speaker. To distinguish among these classes of VSUs, I will refer to “backchannels” and “acknowledgments” interchangeably, and to “other VSUs” and “positive replies” synonymously. Finally, given that previous studies (on Italian Savino 2010, 2011, 2014) report an interaction between intonation, token type, and backchannel turn-taking function, acknowledgements will be further distinguished according to their turn-taking function.

4.2.1 Corpus

Similar to the study on turn-taking contained in Chapter 3, the basis for this backchannel analysis consists of thirty-nine Map-Task dialogues collected in Italian L1, forty in German L2 spoken by the same Italian speakers and nineteen in German L1. Learners’ proficiency levels ranged from A2 to C1 on the CEFR scale. However, for the sake of determining potential effects of proficiency using two balanced groups, they were recategorised into two groups only: beginner (from A1 to B1 levels) and advanced learners (from B2 to C2 levels).²

The resulting corpus includes a total of 2147 VSUs, of which 1745 were BCs and 402 other VSUs. 315 tokens (15% of the extracted data) were excluded from

²See Section 3.2.2 for a description of the task and Section 1.3 for details about participants, data collection, and learner proficiency levels.

prosodic analysis because they did not display the necessary amount of periodic energy to perform a prosodic analysis, e.g. items produced with creaky voice, or items with a voiced portion that was too short, such as *sì*. Accordingly, 1572 BCs and 260 other VSUs underwent prosodic analysis.

Table 4.1 summarises the amount of tokens found for each category, language and proficiency group: total amount of VSUs (the totality of tokens independently of their function), the amount of backchannels (BCs), and the number of positive replies (other VSUs). The entries marked by the abbreviation PA refer to the amount of tokens which underwent prosodic analysis.

Table 4.1: VSU corpus size across language groups.

	IT L1	GL2 (beginner)	GL2 (advanced)	GE L1
Total VSUs	602	273	517	755
BCs	496	223	422	603
Other VSUs	106	50	95	152
Total VSUs (PA)	414	213	427	722
BCs (PA)	389	188	368	586
Other VSUs (PA)	25	25	59	136

4.2.2 Procedure and Metrics

All VSUs produced during the dialogues were annotated using Praat (Boersma & Weenink 2021). After token annotation and extraction, the F0 trajectory of the extracted tokens was pre-processed through smoothing and manual correction of pitch points. The analysis of backchannels takes into account several aspects of their realisation, i.e. their frequency, length, lexical type, structure, function, and intonation. Moreover, for the aspects of type, function, and intonation, a comparison between BCs and positive replies is provided. Table 4.2 summarises all aspects of BC and VSU realisation analysed.

Frequency is operationalised as backchannel rate per minute, while *length* is their duration in milliseconds.

Type encompasses both lexical and non-lexical forms. In this corpus, the most frequent lexical types were *ja* and *sì* (the German and Italian equivalents of ‘yes’, respectively), *genau* and *esatto* (meaning ‘exactly’ in German and Italian, respectively), and *okay*. The most common non-lexical type was *mmhm*. These types

Table 4.2: Aspects of BC and other VSU analysed and their operationalisations.

BC aspects	Operationalisation
Frequency	BCs/minute
Length	Duration in ms
Type	Lexical/non-lexical realisation
Structure	Simple
	Repeated
	Complex
Function	Passive reciprocity (PR)
	Incipient speakership (IS)
Intonation	Rising
	Level
	Falling
Other VSU aspects	Operationalisation
Type	Lexical/non-lexical realisation
Function	Reply to tag question
	Reply to yes-no question
Intonation	Rising
	Level
	Falling

constituted 92% of the entire corpus. The category “other” was used for less frequent token types.

Following a classification similar to Tottie (1991), *structure* classifies the complexity of token form into simple, i.e. one single utterance such as *yes*, repeated,³ i.e. repeated simple tokens such as *okay okay*, and complex, i.e. combinations of different tokens, such as *okay yes*.

Backchannels were also categorised based on their *turn-taking function*, specifically passive reciprocity (PR) and incipient speakership (IS) (Savino 2010, 2011, 2014). Tokens that were produced without the speaker taking the floor and sim-

³I do not use the word “double” like Tottie (1991) because simple types also occurred more than two times in succession in the present corpus.

ply as signals to the primary speaker that they may continue, were labelled as acknowledgement tokens marking PR. When a speaker used backchannels to acknowledge the interlocutor's turn but then took the floor by continuing to speak and causing a turn transition, these backchannels were labelled as marking IS. Interestingly, this corpus includes considerably more PR (1376) than IS (368) tokens. With regard to positive replies, the instances found in this corpus can fulfil functions of either an answer to a yes-no or to a tag question.

Finally, intonation was categorised as rising, flat or falling and measured in semitones (ST) with a reference value of 1 Hz. For each token, F0 points were extracted from two time points, one at the beginning and one at the end of the signal. Depending on the location of the first voiced sound, F0 points were sampled at 10%-90%, 20%-80% or 30%-70% of the token duration. Following Wehrle (2023), a rising intonation was defined as a difference greater than +1 ST between the initial and final F0 points, a level intonation was defined as a difference within +/- 1 ST, and a falling intonation was defined as a difference less than -1 ST.

4.3 Results

4.3.1 BC frequency

Figure 4.1 illustrates the frequency of BCs per minute of dialogue across different groups. The graph reveals a similar rate of BC use among native speakers of German (5.82 BCs per minute) and Italian (5.14 BCs per minute). In contrast, learners exhibit a lower BC rate than that of both native groups. Beginner learners, in particular, produced the fewest BCs, at nearly half the rate of their target language (2.87 BCs per minute). Advanced learners, instead, show a BC rate closer to that of the native German speakers (4.75 BCs per minute). This result may lead to the conclusion that, as learners' proficiency and fluency improve, their backchanneling behaviour tends to approach the native German target. However, this observation is incomplete, as revealed by a more detailed analysis of the individual dyadic interactions.

Figure 4.2 shows the crucial influence of dyad-specific behavior across all groups. Learners display a remarkably similar BC rate across L1 and L2. For instance, dyad BS presents nearly identical rate values in both languages (as shown by the overlapping squares). Moreover, the low BC rate in the beginner group is partly due to the peculiar behaviour of beginner dyad GS. Furthermore, the low backchannel rate observed in the beginner group can be partially attributed to the unique behaviour of dyad GS. The extremely low backchannel production in their L2 output is likely not solely due to their limited German profi-

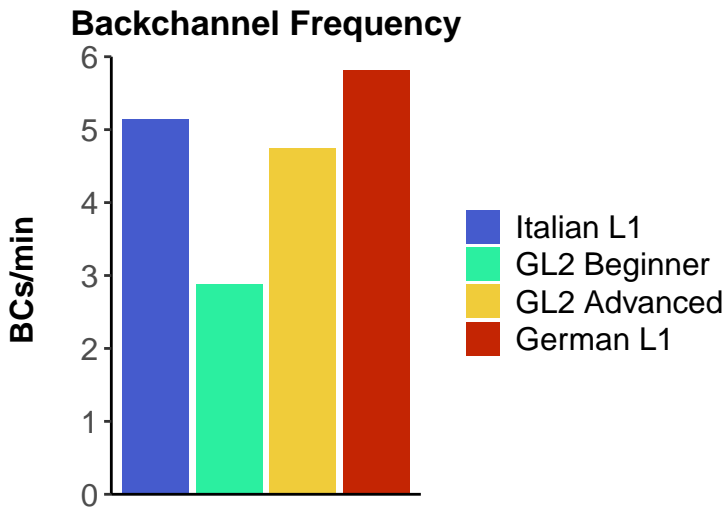


Figure 4.1: Backchannel frequency operationalised as rate per minute of dialogue. The number of BCs per minute is displayed on the y-axis. Language groups are shown on the x-axis and are colour-coded: blue for Italian learners' native speech; aquamarine for beginner learners in L2 German; yellow for advanced learners in L2 German and red for the native German control group.

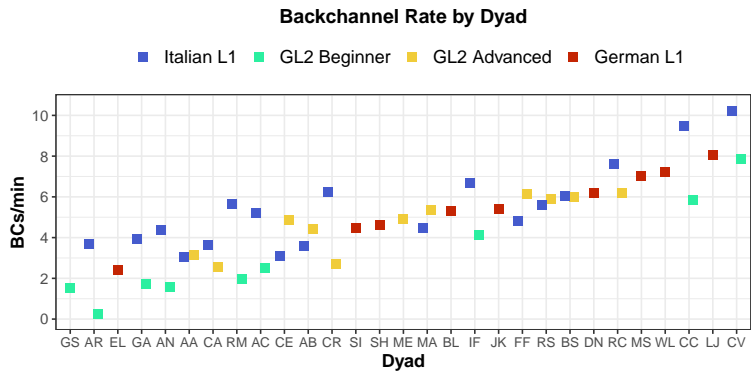


Figure 4.2: Backchannel frequency by dyad operationalised as rate per minute of dialogue. The number of BCs per minute is displayed on the y-axis. Dyads are shown on the x-axis and language group is colour-coded: blue for Italian learners' native speech; aquamarine for beginner learners in L2 German; yellow for advanced learners in L2 German and red for the native German control group. Italian learners of L2 German present two values corresponding to their L1 and L2 speech, distinguished by the colour of the square.

ciency, as they also produced no backchannels in their L1 (and only very few VSUs, which are not displayed in the graph). This suggests that their behaviour is dyad-specific. On the other extreme end, another beginner dyad, CV, presents the highest BC rate across all groups, which would not be expected from group-level results. This high degree of dyad variability is also evident within the native German speaker group. Notably, dyad EL's rate is very similar to that of the beginner dyad GS, indicating that very low backchannel frequency can also occur among native German speakers. This observation challenges the notion of a specific target backchannel rate for learners to achieve. Instead, it suggests that backchannel frequency is largely dependent on the specific dynamics of each dyadic interaction.

4.3.2 BC length

Figure 4.3 shows BC length, i.e. their duration in milliseconds (ms). Here, a difference emerges between the two native language groups: native Italian speakers produce longer BCs (475 ms, SD = 177 ms) than native German speakers (329 ms, SD = 133), suggesting a potential target for learners. The two learner groups show similar backchannel durations with no apparent effect of proficiency, falling between the values found for the two native languages (Beginners: 405 ms, SD = 155; Advanced: 409 ms, SD = 140). At first sight, this might suggest that during the learning process learners tend to approach the target, but eventually plateau. However, this interpretation is again incomplete, as demonstrated by an analysis of the individual dyadic interactions.

Similarly to frequency, by-dyad values for length displayed in Figure 4.4 show that dyad-specific behaviour yields a better explanation for the results observed at least for half of the learner dyads. Indeed, especially advanced dyads (RC, CA, AA, RS, BS, AB, CR), but also some beginner ones (IF, CC, CV), present very similar length values across their L1 and L2. Moreover, it seems that the trend of reducing BC length in the L2 is especially present in dyads with very high length values in their L1 (around and above 500 ms), which is more often the case in beginner (AN, RM, GA, AC) than in advanced learners (CE, FF, MA). This might suggest that learners do perceive a difference in BC length across Italian and German and tend to shorten BCs in their L2, particularly when they perceive their native Italian backchannel length to be highly different from what they categorise as native German. However, this hypothesis is a mere speculation and should be tested on a larger dataset.

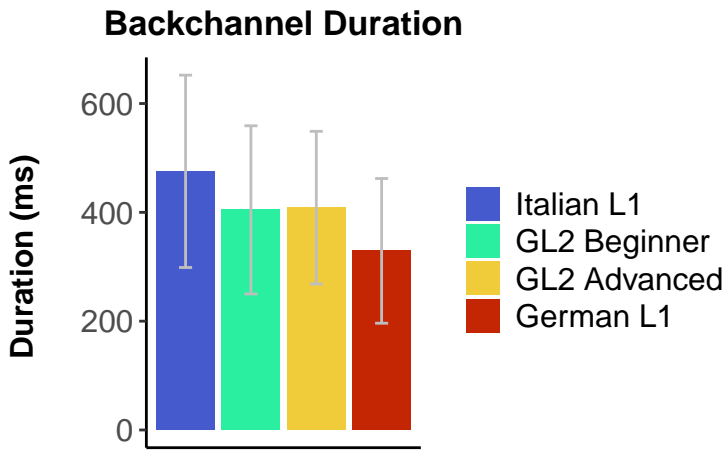


Figure 4.3: Backchannel length operationalised as their duration in ms. Milliseconds are displayed on the y-axis. Language groups are shown on the x-axis and are colour-coded: blue for Italian learners’ native speech; aquamarine for beginner learners in L2 German; yellow for advanced learners in L2 German and red for the native German control group. Gray lines represent the standard error.

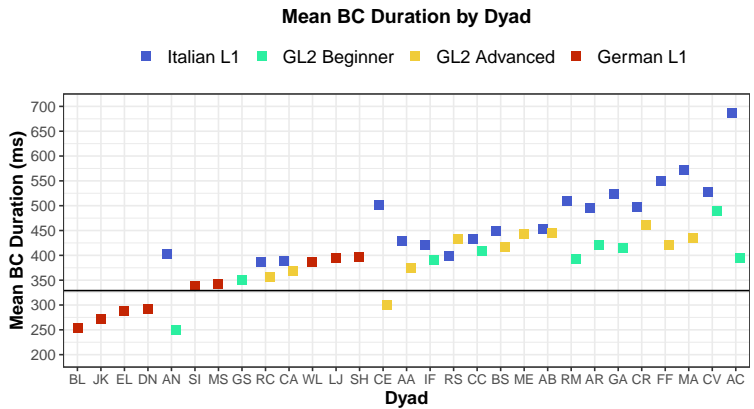


Figure 4.4: Backchannel length by dyad operationalised as their duration in ms. Mean BC duration in milliseconds is displayed on the y-axis. Dyads are shown on the x-axis and language group is colour-coded: blue for Italian learners’ native speech; aquamarine for beginner learners in L2 German; yellow for advanced learners in L2 German and red for the native German control group. Italian learners of L2 German present two values corresponding to their L1 and L2 speech, distinguished by the colour of the square. The horizontal black line corresponds to the mean BC duration of the L1 German group pooled across all speakers.

4.3.3 BC structure

Figure 4.5 shows the percentages of different BC structures across groups, enabling an exploration of the potential relationship between BC structure and length. Specifically, the goal is to check whether the higher proportion of repeated and complex BCs contributes to longer BCs in L1 Italian speakers as compared to L1 German speakers.

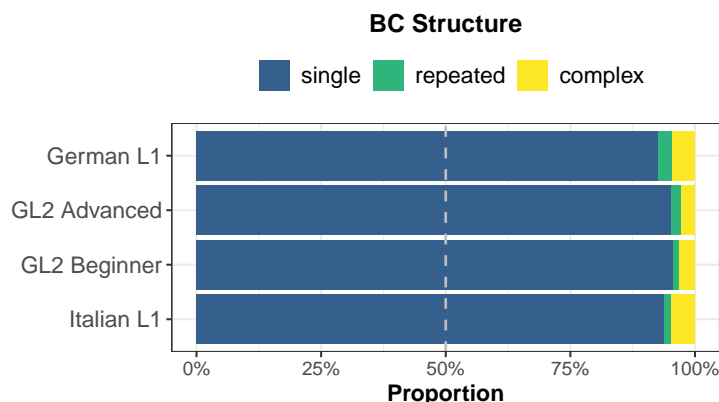


Figure 4.5: Backchannel structure. Proportions of BC structures are shown in percentages on the x-axis. Language groups are shown on the y-axis and are each assigned one bar. Different BC structures are listed in the legend and are colour-coded: blue for single, green for repeated and yellow for complex BCs.

Proportions of BC structures are highly similar across all groups, showing no particular tendency of Italians producing more repeated or complex BCs, which could explain the difference in BC length across L1s. On the contrary, German L1 speakers tended to produce slightly more repeated BCs than L1 Italian speakers (Italian L1: single = 93.8%, repeated = 1.41%, complex = 4.83%; German L1: single = 92.5%, repeated = 2.82%, complex = 4.64%).

Figure 4.6 shows that no particular structure exhibits a dramatically different length that would explain the observed variations. For both L1 Italian and L1 German, simple BC length is very similar to the mean value of BC duration, and in both cases repeated and complex BCs are about 200–250 ms longer (Italian L1: single = 458 ms, repeated = 721 ms, complex = 741 ms; German L1: single = 314 ms, repeated = 498 ms, complex = 521 ms). Interestingly, learners shorten their simple BCs but do not proportionally reduce the length of their repeated and complex BCs, which remain similar to those produced in their L1 (Beginners: single = 389 ms, repeated = 720 ms, complex = 731 ms; Advanced: single = 395

ms, repeated = 719 ms, complex = 645 ms), with the exception of complex BCs in the advanced group.

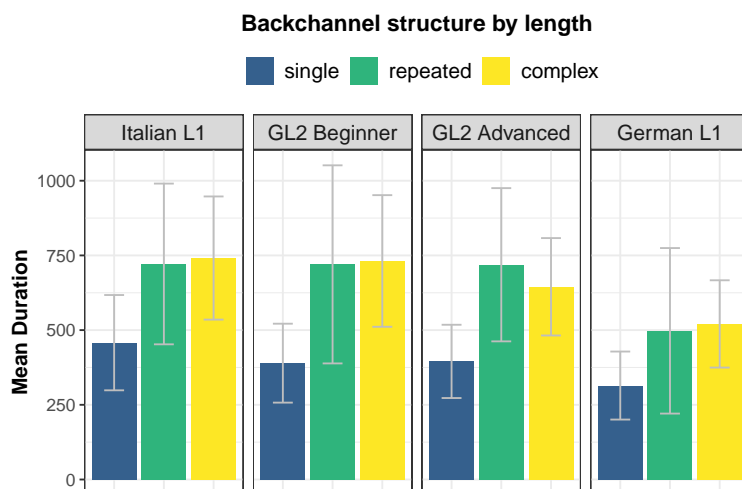


Figure 4.6: Backchannel structure by length for each language group. Language groups are assigned a box each with mean BC length for each type of BC structure. Mean duration of BCs structures are shown on the y-axis. The different BC structures are displayed on the x-axis and are colour-coded: blue for single, green for repeated and yellow for complex BCs. Gray lines represent the standard error.

Finally, the observed length differences cannot be traced back to differences in the syllabic structure of the lexical items used, i.e. to the production of a greater number of syllables. As will be discussed below, the BC types that differ between German and Italian are two monosyllabic types for ‘yes’ (*ja* and *sì*, respectively), and ‘exactly’, which is disyllabic in German (*genau*) and trisyllabic in Italian (*esatto*). In turn, the Italian equivalent is used less frequently than the German one. Even excluding the category “other” did not alter the results in a relevant way. This suggests that measures such as speech or articulation rate might explain the differences in BC length between the languages.

4.3.4 BC type

Figure 4.7 illustrates the proportions of BC types across groups. A comparison of the two native language groups reveals a divergence in preferred BC types. While both groups use *mmhm* in similar proportions (37% in L1 German and 28% in L1 Italian), L1 Italian speakers prefer *okay* (43%) over *sì* (23%), whereas L1

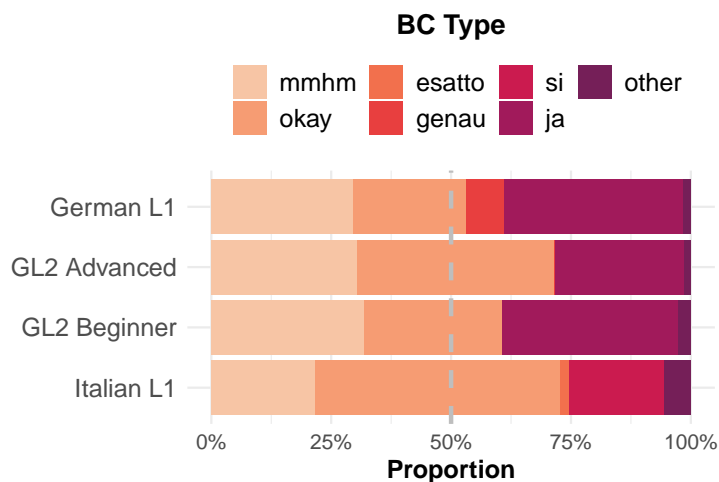


Figure 4.7: Backchannel types. Proportions of BC types are shown in percentages on the x-axis. Language groups are shown on the y-axis and are each assigned a bar. The most-used BC types are listed in the legend and are colour-coded. The category “other” refers to types that rarely occurred.

German speakers show the opposite preference (*ja* 36%, *okay* 20%). Both groups use *genau* and *esatto* infrequently (5% in L1 German and 2% in L1 Italian).

Learners show varying proportions of BC types. Beginners more closely resemble their target language than advanced learners (beginners: *ja* 37%, *mmhm* 32%, *okay* 29%; advanced: *ja* 27%, *mmhm* 30%, *okay* 41%). However, this result warrants further investigation to determine the influence of dyad-specific preferences. Additionally, L2 learners may be transferring their L1 BC type preferences, as the type *genau* is completely absent in the beginners’ data and constitutes only 0.47% of BC occurrences in the advanced learner group. A possible explanation is that its Italian equivalent is rarely used, suggesting that learners require more L2 experience and exposure to begin using this type of BC when speaking German.

Figure 4.8,⁴ which displays the choice of BC type by dyad, reveals a tendency for L1 dyad-specific patterns to be replicated in the L2, particularly among advanced learners (compare L1 and L2 in IF, CV, AN for beginner and AB, RS, CR, CA, BS, AA for advanced learners). This observation could be explained by individual preferences. However, it is also possible that, as learners become more comfortable in the L2, they unconsciously approach their own L1 spontaneous

⁴Remember that dyad GS did not produce any BCs in their L1 Italian (and only one VSU), while the Italian L1 file for dyad ME turned out to be damaged and was therefore not analysable.

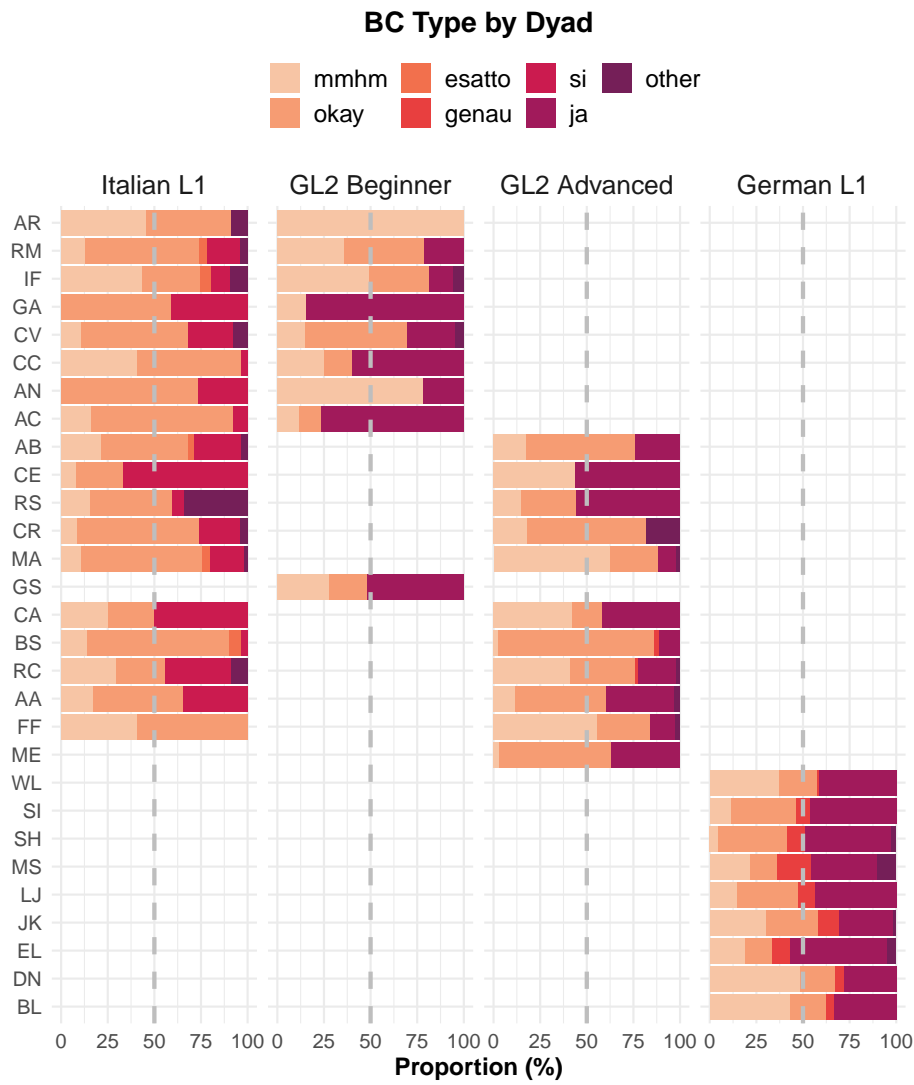


Figure 4.8: Backchannel type by dyad. Proportions of BC types are shown as percentages on the x-axis. Dyads arranged by language group are shown on the y-axis and are assigned one bar each. The most frequently used BC types are listed in the legend and are colour-coded. The category “other” refers to types that rarely occur.

speech style. Essentially, with greater proficiency, L2 speech patterns seem to stabilise, whereas less proficient learners show more variability (independently of the model pattern being that of the native or the target language).

Finally, an evident difference between L1 Italian and L1 German concerns the proportions across types in a by-dyad comparison. It seems that the choice of BC types within L1 German is more consistent across dyads, whereas it is more variable and dyad-dependent in L1 Italian.

4.3.5 BC type by function

Figure 4.9 shows how the proportions of BC types vary depending on their function: passive reciprocity (PR, non-turn-initiating) or incipient speakership (IS, turn-initiating).

For PR, L1 German and L1 Italian speakers show a similar behaviour, primarily using *okay*, *ja/sì* and *mmhm*. Both L1s display roughly the same amount of *mmhm* (25% and 22%, respectively), but L1 German speakers prefer *ja* (43%) over *okay* (22%), while L1 Italian speakers show the opposite pattern, with *okay* (41%) being preferred over *sì* (23%). In both languages, *genau* (6%) and *esatto* (2%) are rarely used for this function.

However, the two language groups diverge considerably in their preferred backchannel types for IS. Italian speakers almost exclusively use *okay* (76%), while German speakers use *genau* (20%) more frequently than its Italian counterpart *esatto* (2%). Interestingly, the non-lexical *mmhm*, frequently used for PR, is only occasionally used for IS in both languages (7% in German and 3% in Italian). Its non-lexical nature likely leads speakers to perceive *mmhm* as non-intrusive, making it unsuitable for signaling an intention to take a turn. Instead, it encourages the other speaker to continue, potentially characterising *mmhm* as a prototypical continuer.

Italian learners of German appear to transfer their L1 PR backchannel preferences to their L2, with the exception of two instances of *genau* produced by advanced learners. For IS, advanced learners use *ja* more often than its Italian equivalent *sì* (29%), but their most frequent choice remains *okay* (65%), mirroring their L1 pattern. Notably, none of the learners use the most typical German type, *genau*, for this function. Beginners produce too few instances of IS backchannels (17 items only) to allow for firm conclusions about their type choices to be drawn. Their limited L2 proficiency may lead them to avoid actively taking turns, preferring to let their interlocutor lead the interaction.

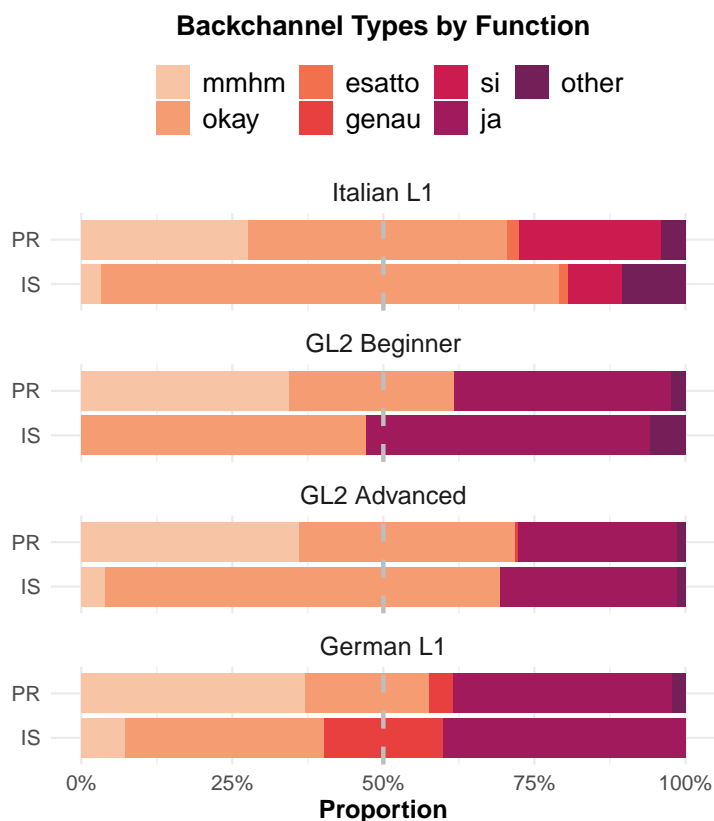


Figure 4.9: Backchannel types by function. Proportions of BC types are shown in percentages on the x-axis. Functions are shown on the y-axis and are assigned one bar each: PR for passive reciprocity and IS for incipient speakership. The most-used BC types are listed in the legend and are colour-coded. The category “other” refers to types that rarely occurred.

4.3.6 Other VSU type by function

Figure 4.10 illustrates the choice of BCs (acknowledgements) and other VSUs (replies to yes-no and tag questions) by function and across language groups. The bars representing acknowledgments correspond to those in Figure 4.7 and are repeated here allowing for a direct comparison.

The two native languages vary greatly regarding the distribution of types across the two reply types. In response to yes-no questions, Italians predominantly use *si* (80%), a preference mirrored by both beginner and advanced learners (96% for both groups). German speakers, however, use a wider range of re-

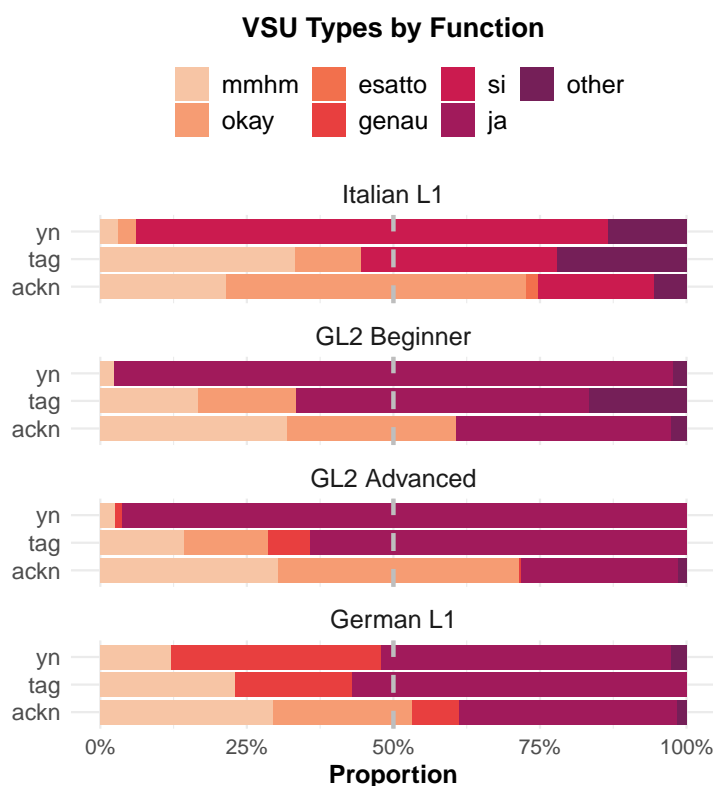


Figure 4.10: Very short utterance types by class across language groups. Proportions of VSU types are shown in percentages on the x-axis. Classes are shown on the y-axis and are assigned one bar each: replies to yes-no questions, replies to tag questions and acknowledgements (BCs, for comparison). The most-used VSU types are listed in the legend and are colour-coded. The category “other” refers to types that rarely occurred.

sponses, with only a few instances of *mmhm* (12%), more predominant *ja* (50%) and many *genau* (36%) items. For tag replies, Italians seem to equally prefer *sì* and *mmhm* (33% each), with occasional use of *okay* (11%). German speakers primarily use *ja* (57%), followed by *mmhm* and *genau* (23% and 20%, respectively). L2 learners reproduce the preference for *ja* (beginners: 50%; advanced: 64%), but their overall pattern is closer to their native Italian, with some occurrences of *okay* and almost no use of *genau*, which appears only once in each of the two reply classes among advanced learners. The type *genau* appears to be a hallmark of L1 German, used in both yes-no and tag replies (36% and 20%, respectively), whereas this is not the case in L1 Italian and the interlanguage.

Finally, comparing the two replies to acknowledgements, it is evident that the choice of type changes in terms of proportions across classes and functions, suggesting a relation between type choice and function expressed.

4.3.7 BC intonation

In this section, BC intonation contours are explored in relation to their function and type. Figure 4.11 shows a tendency for PR backchannels to be expressed with rising intonation and IS backchannels with falling intonation, consistent

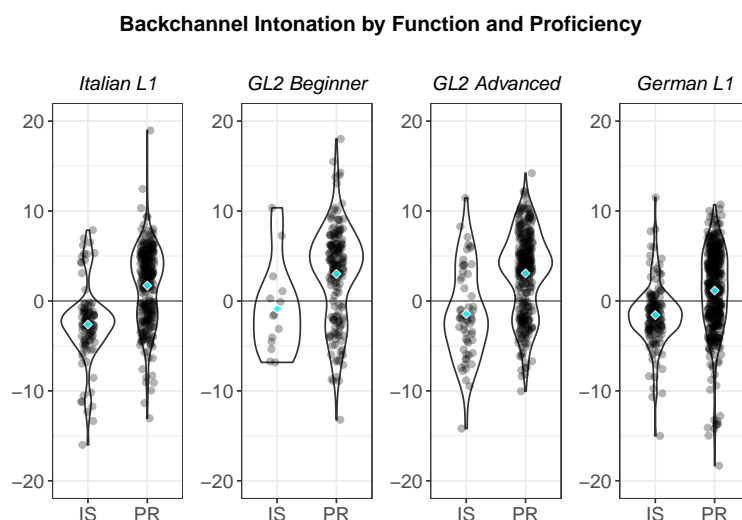


Figure 4.11: BC contours by function across language groups. Values above zero represent items with a falling contour; values below zero represent items with a rising contour. Cyan diamonds represent mean values.

with previous results (Savino 2010, 2011, 2014; Wehrle 2023). This pattern holds across all language groups under investigation, but distributions of values suggest some differences. Italian learners of German, both in their native Italian and their L2 German, seem to avoid flat intonation contours (values around zero), unlike native German speakers. Moreover, L2 German shows a considerable influence from native Italian patterns, but with higher variability, as is typical of an interlanguage. Finally, a small proportion of PR backchannels is expressed with a falling contour in all language groups.

To assess whether this latter result can be explained by other variables, Figure 4.12 and Figure 4.13 display BC intonation contours by function and lexical type in a continuous and categorical fashion, respectively, so that proportions can be related to the amount of data for each type and function. It becomes clear that this apparent relationship between contour and pragmatic function is more nuanced, as intonation is also dependent on word choice. Indeed, two types, *mmhm* and *genau*, exhibit preferred contours regardless of function. *Mmhm* is

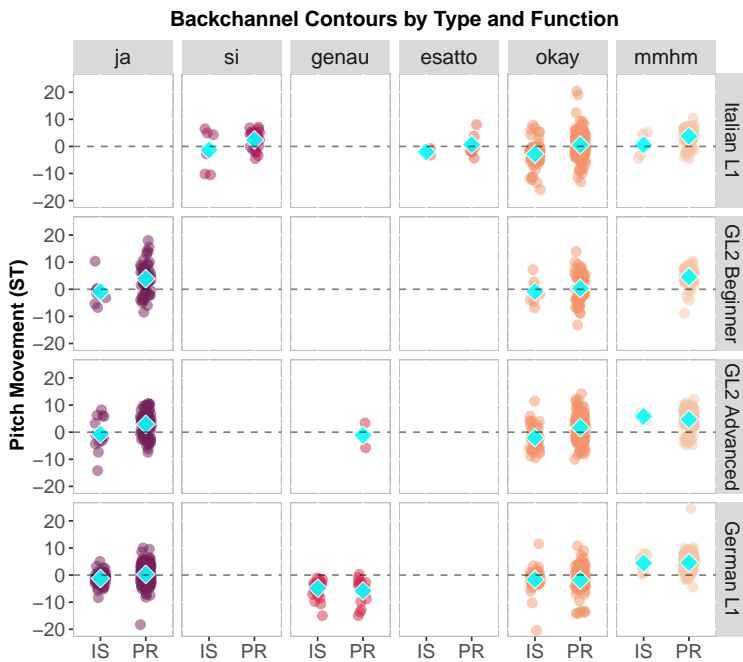


Figure 4.12: BC contours by type and function across language groups. Values above zero represent items with a falling contour; values below zero represent items with a rising contour. Cyan diamonds represent mean values.

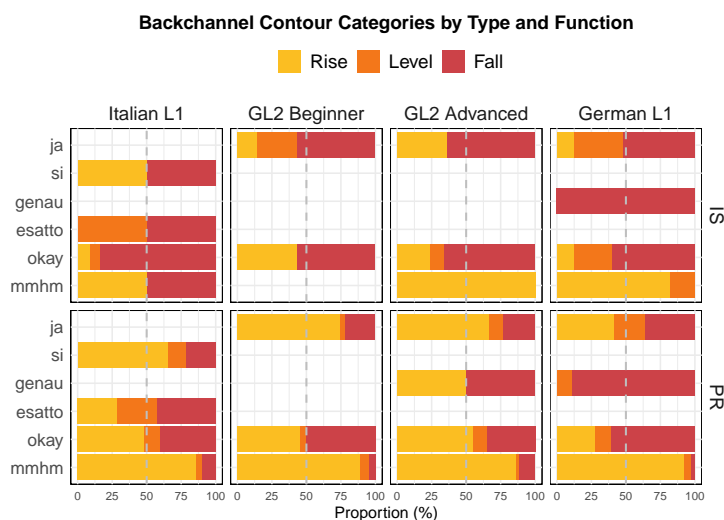


Figure 4.13: Categorical classification of BC contours by type and function across language groups. Proportions of BC contour categories are shown in percentages on the x-axis. BC types are displayed on the left of the y-axis and are each assigned a bar. Upper boxes refer to types used with an incipient speakership function (IS), while bottom boxes refer to types used with a passive reciprocity function (PR).

typically rising across all language groups, while *genau* is predominantly falling in L1 German (with 11% level in PR) and never rising. In contrast, the Italian equivalent *esatto* shows similar proportions of rising (29%), falling (43%) and level contours (29%) in the PR condition, indicating no specific contour association with this word in Italian. Learners transfer this variability to the corresponding *genau* in their L2, using both rising and falling contours. However, these observations are based on limited data for *esatto* in L1 Italian (seven PR tokens) and *genau* in L2 German (two PR tokens produced by advanced learners, one rising, one falling). As previously noted, *esatto* is not common in L1 or L2 Italian, unlike the more frequent use of *genau* in L1 German.

The lexical type *okay* shows a similar distribution of rising and falling contours for PR in native Italian (41% falling, 11% level and 48% rising contours) and in L2 German (beginners: 51% falling, 3% level and 46% rising; advanced: 35% falling, 10% level and 55% rising). Native German speakers, however, prefer falling contours for *okay* even when used with a PR function (61% falling, 12% level and 27% rising). This variability in contours for *okay* used with a PR function might play the biggest role in explaining the broad range of values observed for PR in Figure 4.11, across groups. When expressing IS, *okay* instead tends to show falling

contours across all language groups (L1 Italian: 84%; L2 beginner learners: 57%; L2 advanced learners: 67%; L1 German: 60% falls and 28% levels).

Finally, *ja* and *sì* best illustrate the contour-function relation shown in Figure 4.11 for Italian speakers. When expressing PR, Italian speakers predominantly use *sì* with a rising contour in both their L1 and L2 (65% in Italian; 74% and 67% in L2 German by beginner and advanced learners respectively), while native German speakers still prefer falling contours (61%), similar to the case of *okay*. When expressing IS, L1 German aligns with the general trend of using falling and level contours for this function (60% and 28% respectively). In L1 Italian, the limited number of instances (six tokens, three rising, three falling) prevents firm conclusions, and more tokens may yield different results. As reported in Section 4.2.1, a prosodic analysis of many *sì* tokens was not possible due to the shortness of their voiced portion. However, L2 German has more *ja* tokens (130 for PR and 21 for IS across proficiency groups), and given learners' tendency to mostly transfer their L1 intonation patterns, it could be hypothesised that their L2 mirrors the L1 also in the case of *sì* used with an IS function. This would support the observation that Italian speakers tend to use a rising contour for *sì* in PR (74% rises and 3% levels by beginner learners; 67% rises and 7% levels by advanced learners) and a falling contour in IS (57% falls and 29% levels by beginner learners; 64% falls and 36% levels by advanced learners).

Overall, learners' patterns of contour-type-function relations closely resemble those of their native language, indicating a transfer from their L1. Moreover, proficiency level appears to have no relevant impact, which is particularly evident in the categorical analysis of the PR function (Figure 4.13), where more data points yield more reliable results.

4.3.8 Other VSU intonation

A final observation is related to other classes of VSUs using the same lexical types. Figure 4.14 shows contours of types used as positive replies to yes-no and tag questions in Italian L1, German L2 and German L1. Interlanguage data is combined across proficiency levels, since no major differences were observed in the analysis reported in the previous section. As stated above, the Italian *sì* used for PR has both falling and rising contours in equal proportion. However, in yes-no replies, *sì* is predominantly falling in Italian (86% of 15 items). L1 German also primarily uses falling contours for both tag and yes-no replies (53% for both, based on 17 tag and 51 yes-no tokens). Moreover, the tendency for *mmhm* to have rising and *genau* to have falling contours is consistent across all groups. Further analysis on positive replies is not possible, since the amount of data for these

token classes is very limited and does not allow for reliable observations. Nevertheless, this preliminary view provides further evidence for the observation that the function of some lexical types influences their prosodic realisation.

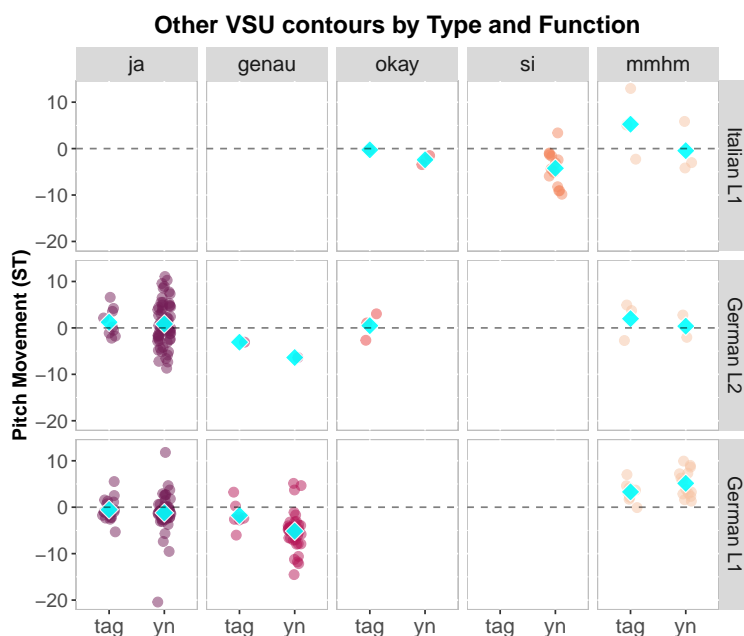


Figure 4.14: Other VSU contours by type and function across language groups. Values above zero represent items with a falling contour; values below zero represent items with a rising contour. Cyan diamonds represent mean values. Functions of the replies are shown on the y-axis: replies to yes-no questions (yn) and replies to tag questions (tag).

4.4 Conclusion

This contribution offered an in-depth analysis of BCs and other VSUs in Italian and German, as well as in the L2 German spoken by Italian learners. The analysis covered BC frequency, duration, structure, lexical (and non-lexical) type, function, and prosodic realisation. Other VSUs sharing lexical types with BCs but serving different functions (e.g. positive replies to yes-no and tag questions) were also included for comparative purposes. Dyad-specific variability was considered because: 1) individual conversational behaviour depends not only on idiosyncratic factors and speakers-specific speech style but also, crucially, on the unique dynamics that arise from the interaction between the two specific parties

in the conversation; and 2) this helps distinguish genuine acquisition or transfer of patterns from dyad-specific behaviour.

It was found that German and Italian are quite similar regarding BC frequency. In L2, analysing only group-level data could have falsely suggested that target-like patterns of BC frequency are achieved along with increasing proficiency. Instead, a by-dyad analysis revealed similar behaviour in learners' L1 and L2, suggesting that dyad-specific patterns are more important than proficiency levels when examining the rate of BCs produced. Moreover, the by-dyad variability observed even within the group of native German speakers challenges the notion of a fixed target frequency for learners to acquire.

Cross-linguistic differences exist for BC duration. Italian speakers produce longer BCs than German speakers, a difference not explained by an analysis of BC structure. Learners' BC durations fall between those of the two native languages. However, in this case, too, dyad-specific behaviour better explains learners' L2 patterns in at least half of the dyads, and proficiency appears to have no impact. Intriguingly, Italian dyads with very long BC durations were the same ones that tended to shorten them in their L2. Possibly, a larger perceived difference between native and target language norms might encourage an adaptation to the target. This hypothesis requires further investigation.

Regarding lexical choice, distinct language-specific relationships between type and function were observed, suggesting this could be a learning target for L2 learners. It was found that learners tend to favour BC types shared with their native Italian over German-specific ones like *genau*. A by-dyad analysis revealed that advanced learners, in particular, tend to use BC types in proportions similar to their L1. One possible explanation is that higher proficiency allows learners to transfer their L1 spontaneous speaking style to the L2, resulting in a more consistent output compared to the highly variable production of beginners. Only advanced learners used German-specific BCs, albeit infrequently, indicating a positive effect of increased target language exposure. Moreover, beginners produced very few BCs with the function of actively taking a turn, possibly due to their lower proficiency and consequent preference to let their interlocutor lead the conversation.

An even more complex, non-arbitrary mapping between lexical type, function, and intonation was observed in both languages. Overall, PR acknowledgements tended to be produced with rising contours, and IS acknowledgements with falling contours, across all groups. However, when BC type was taken into account, it emerged that this is not a one-to-one relation. A possible example of this function-contour relation is the case of *ja* and *sì*, but limited L1 Italian data prevented a reliable analysis of this trend. Hypothesising that learners transfer

their L1 patterns to the L2, it could be assumed that this tendency exists in L1 Italian as well, confirming that the prosody of some types is influenced by their function. On the other hand, the intonation contour was found to be highly variable for certain types, as in the case of *okay*, which presented both rising and falling contours for PR. The types *mmhm* and *genau* exhibited mostly unidirectional, type-specific prosody regardless of function. Other VSUs, i.e. positive replies to yes-no and tag questions, provided further evidence for the influence of function and/or type on the prosodic realisation.

This study used peer interactions, matching learners with learners and natives with natives. Hence, it is only possible to speculate on Italian learners' backchannel use in a conversation with native German speakers. An interesting direction for future research therefore involves studying mixed dyads to evaluate the potential negative effects of learners' differing BC use on the success and perception of the communicative exchange. Larger corpora are needed to confirm the robustness of the observed trends, since this study's limited sample of token types and functions did not allow for reliable statistical testing. Finally, future research on backchannel intonation could provide a more fine-grained analysis, such as by using periodic energy measures. These metrics, being based on periodic cycles roughly corresponding to syllables, would require experimenters to make methodological decisions to define backchannel structure in greater detail, since the overlap between syllables and backchannels can be less stable than expected in such short utterances.

Despite these limitations, this study offers some fundamental suggestions for further investigations. First, preferential co-occurrences appear to exist between different aspects of BCs (e.g., lexical type, function, and intonation), warranting further investigation of these relationships. Second, these results suggest that dyad-specific patterns seem more predictive of some aspects of L2 backchannel production than proficiency. Therefore, using learners' L1 as a baseline and examining dyad-specific behaviour is important to distinguish individual variability from the transfer or acquisition of patterns. Finally, consistent with the literature, this study found both cross-linguistic similarities in BC use and language-specific aspects that are not correctly reproduced in the L2. This suggests an incomplete acquisition of target-like backchannelling behaviour in the L2. Therefore, comparative studies of diverse language pairs, like the present one, can raise awareness of culture- and language-specific conventions. These findings should be addressed in L2 pedagogy in order to improve learners' intercultural communication skills.

5 Conclusion

The goal of this book was to provide a theoretical and methodological foundation for understanding learners' development of spoken interactional skills in the classroom context, with the ultimate aim of advancing the applied field of second language teaching. Its motivation was driven by the principles of the Common European Framework of Reference for Languages (CEFR, Council of Europe 2001), which describes learners as “social agents”, who constantly need to accomplish communicative tasks as members of society. Hence, language competence is considered as communicative competence, emphasising its interactional aspects rather than its mere linguistic aspects, such as grammar and the lexicon.

Considering the centrality of spoken interaction for daily communication and the descriptors of learners' skills in the CEFR, I carried out three studies on crucial but neglected abilities in second language teaching and learning: prosodic competence – specifically, prosodic highlighting of important or new information in the message – and interactional competence – specifically, two aspects of interactional fluency: turn-taking and vocal feedback signals (backchannels). These studies were carried out on Italian learners of German as compared to German native speakers. However, the results have broader theoretical and methodological implications for second language acquisition, showing that much joint work by researchers and pedagogues is still needed to enable concrete and beneficial applications for L2 learners, irrespective of the native and target languages.

After summarising the relevance and findings of the three studies,¹ I will discuss their implications for second language acquisition research and language teaching, and, finally, indicate some limitations and suggest future directions.

¹The following summaries offer a brief and conclusive overview of the studies conducted. Detailed recapitulations and discussions are available in Section 2.9 for prosodic marking of information status, Section 3.6 for turn-taking, and Section 4.4 for backchannels.

5.1 Summary

5.1.1 Prosodic marking of information status

In Chapter 2, I investigated Italian learners' ability to prosodically mark information status within noun phrases in L2 German. Previous studies report that Italian speakers do not seem to mark post-focal given information in noun phrases, which is instead always accented, and that they transfer this prosodic behaviour to their L2 German. These previous studies have also supported their results with a perception experiment in which Italian listeners could not reconstruct the context of noun phrases from their prosodic realisations only, suggesting that no prosodic cues were available for interpreting the information status of the elicited noun phrases. However, these studies approached the question in terms of a categorical presence or absence of accentuation, overlooking continuous information relating to the modulation of prosodic cues.

I investigated such modulation using an innovative method based on periodic energy and F0 measurements to quantify prosodic aspects related to prosodic strength and F0 contours. Furthermore, I tried to overcome some limitations of previous studies regarding sample size and elicitation method by collecting a larger data sample with an interactive elicitation game.

Results stand in contrast to previous findings with regard to both Italian learners' L1 and L2. The Italian learners of the present study marked post-focal information within noun phrases in their L1 but did so by using distinct F0 modulations on the first word of the noun phrase, instead of prosodically attenuating the post-focal given element, as in West-Germanic languages. Although they tend to transfer their native intonation contour when speaking German, Italian learners are able to reproduce the typically German post-focal reduction of prosodic strength. However, learners apply it in all pragmatic conditions across all proficiency levels (including beginners), suggesting that they might identify this cue to deaccentuation as, instead, a salient marker of native German. Its application irrespective of function can be interpreted either as a form of hypercorrection or as negative transfer from their native language, in which prosodic strength does not mark information status contrasts. These findings were confirmed by an additional analysis from an Autosegmental-Metrical perspective, using established measurements for the continuous investigation of F0 contours and prosodic strength, and a categorical interpretation of the results in terms of pitch accent type.

The current study revealed L1 and L2 prosodic patterns that were not apparent in previous analyses and made a step toward greater ecological validity by

employing an innovative periodic-energy-based method for phonetic analysis and an interactional data collection approach adapted to the need for controlled data. Its results are relevant to L2 teaching, as they identify critical aspects of L2 prosody acquisition that should be addressed in pedagogical contexts.

5.1.2 Turn-taking

In Chapter 3, I discussed the absence of a standardised instrument for the quantification of interactional competence in L2 and proposed a workflow as a possible starting point for an objective assessment of L2 interactional competence. The proposed method includes quantification and visualisation tools of the L2 based on temporal measures. Its informativeness with regard to interaction management in an L2 across different proficiency levels was tested on goal-oriented cooperative dialogues performed by learner pairs matched by proficiency.

Specifically, I adapted a visualisation tool to display the dynamics of floor management in dyadic interactions operationalised as proportions of conversational activities: the time spent speaking for each interlocutor, the total amount of silence, the time of overlapping turns and backchannels (to distinguish the special status of the latter, as they do not constitute turns in themselves). In addition, the total duration of the dialogue was also taken into account. The extracted data were used to assess differences in the proportion of these metrics across proficiency levels and explore which were the best predictors of learner proficiency in terms of similarity to their native baseline.

Overall, the results suggested that silence and speech time of the instruction giver in the task robustly distinguish beginner and advanced learners and that higher proficiency corresponds with less overall silence and more speech time. Conversely, at low levels of proficiency, the cognitive difficulties of speaking an L2 can lead to less fluent interactions, with half of the conversation consisting of silence.

A qualitative by-dyad analysis integrates the by-group results, suggesting that from B2 level, learners present more similar patterns across their L1 and L2; this is an observation which has to be tested further on a larger data sample. Finally, some cross-linguistic/cross-cultural differences in the total duration of the dialogue and speech time of the instruction follower emerged from an exploratory analysis of native German interactions as compared to native Italian ones. This preliminary result suggests that there was a higher degree of diligence by and collaboration between native German speakers in the completion of the task.

These quantification and visualisation tools, based on temporal turn-taking metrics, demonstrate that reliably identifiable conversational activities are re-

lated to L2 proficiency, opening up possibilities for follow-up research in the field of interactional competence assessment, focused on implementing content-related information.

5.1.3 Backchannels

In Chapter 4, I carried out an in-depth analysis of backchannels across native Italian and German and in L2 German spoken by Italian learners. Previous research on backchannelling across languages and in L2 have been conducted only regarding some aspects of backchannel use and on limited language pairs. In order to provide a fully comprehensive view of the phenomenon in a new language pair, the various aspects of backchannels analysed here were: frequency of occurrence, length, lexical vs non-lexical type, structure (single- or multi-word), turn-taking function and prosodic realisation.

From a cross-linguistic perspective, the two native languages were found to have a similar BC frequency, but different BC length not ascribable to their structure (i.e. how many BCs are concatenated into a single BC production), with Italians producing longer backchannels than Germans. A complex, non-arbitrary, language-specific mapping between lexical type, function and intonation was found in both languages. Overall, there is a preference for producing non-turn-initial acknowledgements with a rising contour and turn-initial acknowledgements with a falling contour. Nevertheless, for some types the function seems to be overridden by the (non-)lexical type, so that the prosodic outcome is mostly independent of the function expressed. For learners, dyad-specific variability disentangled apparent cases of a progressive acquisition of target features, such as BC frequency and length, showing that learners do not approximate an ideal L2 target, but tend to instead approach their own L1 baseline. In the case of BC frequency, learners' closer approximation to their L1 baseline results from higher proficiency in the L2, probably as an effect of improved overall fluency. Learners also tended to reproduce their L1 patterns in mapping lexical type, function, and intonation. Only the advanced learners were found to use typically German backchannels, although they did not always match the corresponding intonation contour.

In sum, L2 speakers showed similar backchanneling behaviour in their native language and in the L2, apart from a reduced frequency compared to both native languages. This transfer of native features to the L2 points to possible challenges in intercultural communication and remains to be explored further.

5.2 Implications for second language research and teaching

The results of these three studies are relevant not only for Italian learners of German. There is a great deal of literature dealing with cross-linguistic differences in the prosodic marking of information status, and a typological categorisation has also been proposed. The study presented in this book has shown that new methods and designs can enrich the existing knowledge, bringing to light overlooked linguistic phenomena and encouraging further research on this topic, especially taking into account the improved ecological validity of the experimental procedure used here and the application of various methodological approaches. Despite findings showing that the L2 learners under study do not operate a complete transfer of prosodic patterns as pointed out by previous studies (Swerts et al. 2002; Avesani et al. 2015; Avesani et al. 2013), these learners still show the interference of the L1 phonological rules. Thus, they do not implicitly acquire the correct foreign prosodic patterns in L2 classrooms, missing the link between form and function. This result is in line with other studies on prosodic marking of information status in L2 reporting cases of phonetic and phonological transfer from the L1, or realisations otherwise deviant from the target norm (see Section 2.1.6). Given that the relation between prosodic realisation (form) and meaning expressed (function) differs across languages and that learners are not able to master it solely from the input, the results presented here point to the necessity of creating pedagogical tools applicable to L2 classroom settings. This aim would require a joint effort by pedagogists and phoneticians, as most previous studies (see Section 2.9 for a detailed discussion on prosodic training techniques) have been conducted in the field of phonetics, lacking a strong pedagogical framework, which makes it difficult to identify practicable pedagogical techniques for teaching prosody in L2 classrooms.

Turn-taking and backchannelling conventions are interactional aspects of communication which are not only language-specific, but also culture-specific. Most studies on intercultural interaction have been conducted on cultures and languages belonging to different continents. However, few studies have been conducted on intercultural communication within Europe, probably assuming that the long-established contact among inhabitants of these countries would reduce differences and favour adaptation. The present results on both turn-taking and backchanneling behaviour suggest some language- and culture-specific interactional conventions, which differ even between geographically close language communities among which there is well-established and long-lasting contact (for the specific case of South Italians and Germans, the historical factor

5 Conclusion

of immigration has also contributed to bringing the two cultures closer together). These differences in interactional features and conversational patterns might seem more subtle than those found between some previously explored language pairs, such as American English–Japanese, Canadian English–Chinese or Vietnamese–German, but they are likely to still be noticeable by listeners and can potentially cause misunderstandings and/or misperceptions. Indeed, listeners' high sensitivity to both backchanneling and turn-timing is widely attested. Thus, research investigating language- and culture-specific interactional behaviour can have beneficial pedagogical applications in the field of intercultural communication, and can increase awareness, comprehension and an acceptance of differences in multicultural societies. The findings of the studies presented here on aspects of interactional competence show that these learners do not achieve a target-like reproduction of interactional cues from the sole exposure to the target language in L2 classrooms, which points to the necessity of specific pedagogical tools for interactional skills. The TBLT framework would lend itself well to the case of interactional abilities, as it has shown positive effects on learners' interactional competence in previous studies (Pérez 2016; Waluyo 2019; Fang et al. 2021; see also Mackey et al. 2016 for the theoretical framework).

Currently, both prosodic and interactional skills do not generally receive enough attention in the field of language teaching. In the past, this was due to a predominantly lexicogrammatical approach and a focus on written exercises in both teaching and assessment settings. Nowadays, even within the relatively recent shift towards the communicative approach in L2 pedagogy, which emphasises the use of language rather than its theoretical knowledge, there is often poor empirical understanding of these aspects, preventing them from being discussed and highlighted explicitly in classrooms. Indeed, realizing the central role of these skills in communication is not straightforward from a non-specialist and naïve perspective, as these abilities might not be seen as primary when compared to syntax and vocabulary in the elaboration of a linguistic message. However, especially given the renewed attention to spoken communication, they play a central role.

Prosody is essential for transmitting linguistic and paralinguistic information, as well as subtle shades of meaning, thus ensuring the correct interpretation of the message. A sufficiently smooth turn-taking system between interlocutors is necessary for avoiding the misinterpretation of long silences or long overlaps, together with possible negative culture-specific interpretations. Finally, backchannels are recognised as having a positive social value in conversation by manifesting attention and interest towards the primary speaker, as well as contributing

to smoother turn transitions by signalling a possible desire to take the floor. Empirical evidence shows that unconscious acquisition only partially takes place in L2 classroom settings, and that it does not provide sufficiently good results. As discussed in the introduction of this book, teachers generally do not receive a training on prosodic and interactional competence, and even if native speakers of the language serve as instructors, they might not have clear intuitions about their unconscious use of the language. It is also important to remember that not all teachers are native speakers and might not be completely aware of subtle uses of prosody and cultural-specific interactional conventions. Finally, in some educational settings, language classes can be large, so that learners tend to exercise their communicative abilities among each other during practice sessions and only rarely speak with the teacher.

Due to these factors, the amount and quality of the input received in L2 classroom settings may not be enough to guarantee the implicit learning of prosodic and interactional target patterns. In our globalised and technologically advanced world, in which communication is enabled in real-time and face-to-face independently of geographical distance, applied L2 research with the aim of increasing cross-linguistic and cross-cultural awareness is highly necessary, as is the development of effective didactic tools for L2 learning. This objective requires joint work from both researchers and teachers, bridging the gap between the two fields.

By highlighting the differences between two languages (German and Italian), the three studies presented in this book provide groundwork for pedagogical tools with a contrastive approach, that is, based on this specific language pair as the native and target languages of the learners. At the same time, individuating strategies of prosodic marking of information status, turn-taking and backchannelling relative to each language, results could also be taken as a starting point for developing training materials for learners of Italian and German with different L1 backgrounds.

5.3 Limitations and future directions

Although the findings of this study are of some significance, there are some limitations to their generalisability.

Firstly, as interlanguages are complex systems to which the existing knowledge contributes, results on these particular learners are to be considered in the light of the participants' specific native and target language. Furthermore, the high regional linguistic variability present in Italy does not necessarily permit a

5 Conclusion

generalisation of the findings to other Italian learners of German, especially with regards to phonology.

Secondly, learners were categorised into two main proficiency groups to allow for more reliable statistical testing, since the sample contained more intermediate than beginner and advanced learners (according to the CEFR classification). To investigate the process of second language acquisition and critically discuss it in relation to the CEFR and its descriptors, a larger sample for each proficiency level is required. In this way, it would be possible to observe if learners fulfil the skill descriptors related to each level and, if not, intervene with apposite and efficient pedagogical tools. Ideally, a replication of these studies based on the CEFR proficiency categories could clarify the nature and degree of the variability in learners' interactional behaviour found among groups. A longitudinal study could provide further evidence for the development of L2 skills within each dyad.

Thirdly, two interactive, task-oriented data collection methods were used: a semi-scripted conversational board game and a Map Task. Despite the effort made to design a more interactive board game as compared to previous elicitation methods, and the inclusion of the more spontaneous Map Task, the ecological validity of the experimental design can still be improved. A challenge for future research lies in finding the best compromise possible between the investigation of fully spontaneous speech and systematic data collection, in order to provide robust observations based on a sample which is more representative of real-life communication and of sufficient external validity. Concordantly, it has to be remembered that eye contact was blocked during recordings in order to foreground communication in the vocal channel, thereby eliminating the analysis, if not the use, of all possible non-verbal signals which might have contributed to the interaction in crucial ways. Therefore, a similar investigation with a multimodal approach would enrich our understanding of real-life, face-to-face conversations by investigating how the different channels complement each other and how they interact (for eye-gaze and vocal feedback, see Spaniol et al. 2023; Sbranna, Savino, et al. 2025; for a multimodal approach in L2 interaction, see Tsunemoto et al. 2022; McDonough et al. 2020).

Finally, the data collected for the analysis presented in this book only included conversations among peers (learners with learners and natives with natives) to reproduce and investigate second language acquisition in a classroom setting. However, a second language is learnt not only for use as a lingua franca in intercultural contexts, i.e. as a common language among non-natives, but also to be able to communicate with native speakers. Thus, mixed dyads of learners and natives should be investigated as well, since L1-L2 and L2-L2 conversations have shown differences in interactional features (e.g., Shibata 2023; Kley et al. 2023;

Jung & Crossley 2024). Such conversations could clarify which deviations from the native norms have the most detrimental effect on communication. This assessment could be based on features causing communication breakdowns and/or ratings of the perceived success of the communication. Findings of such a study would highlight the central aspects on which educational tools should be based, with the ultimate goal of improving learners' communicative skills and the ease of interaction.

Despite these limitations, the findings presented in this book enrich the existent body of knowledge in SLA studies and hopefully raise awareness about the extent of implicit learning in L2 classrooms. Future studies should ideally address the gap between research and pedagogy to benefit learners and improve intercultural communication.

Appendix

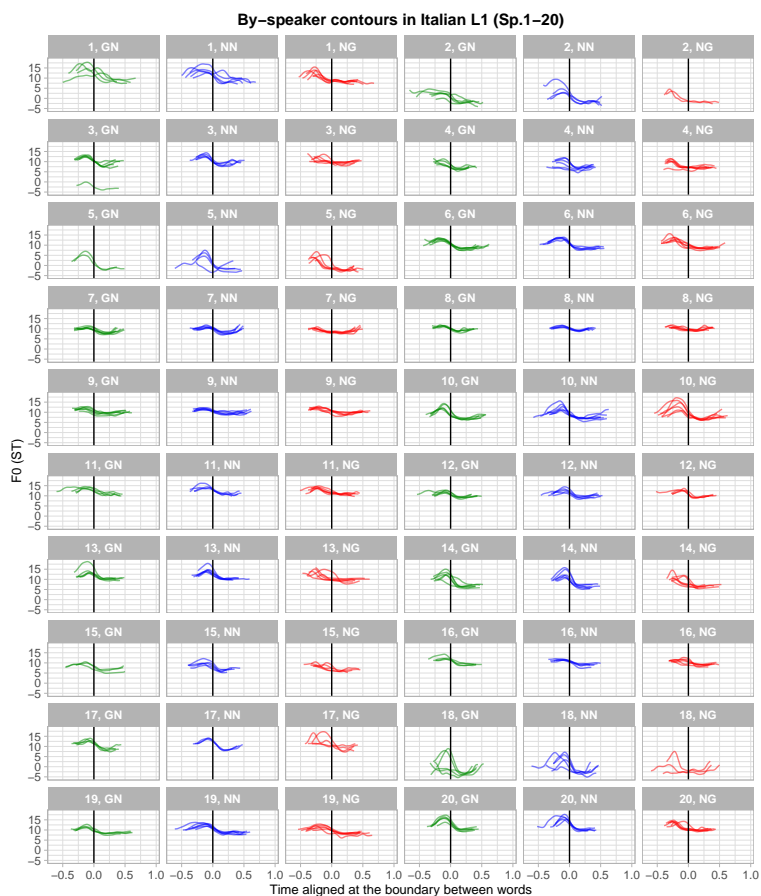


Figure A1: By-speaker F0 contours for the three information structure conditions in L1 Italian. Information structure conditions are colour-coded: green for given-new (GN), blue for new-new (NN) and red for new-given (NG). The black line marks the boundary between words. Speakers are identified by numbers from 1 to 20.

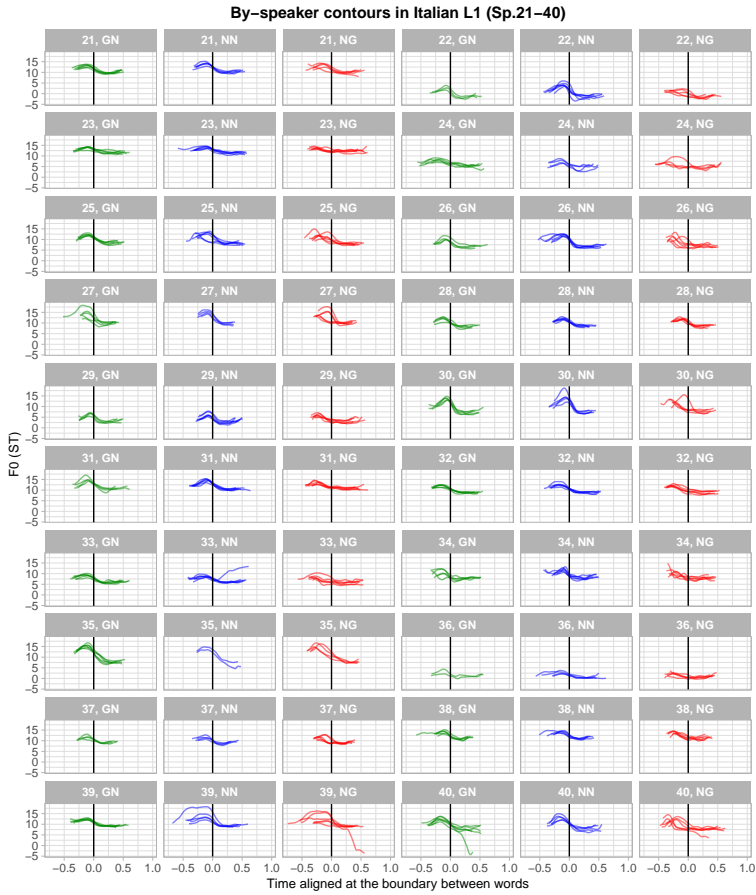


Figure A2: By-speaker F0 contours for the three information structure conditions in L1 Italian. Information structure conditions are colour-coded: green for given-new (GN), blue for new-new (NN) and red for new-given (NG). The black line marks the boundary between words. Speakers are identified by numbers from 21 to 40.

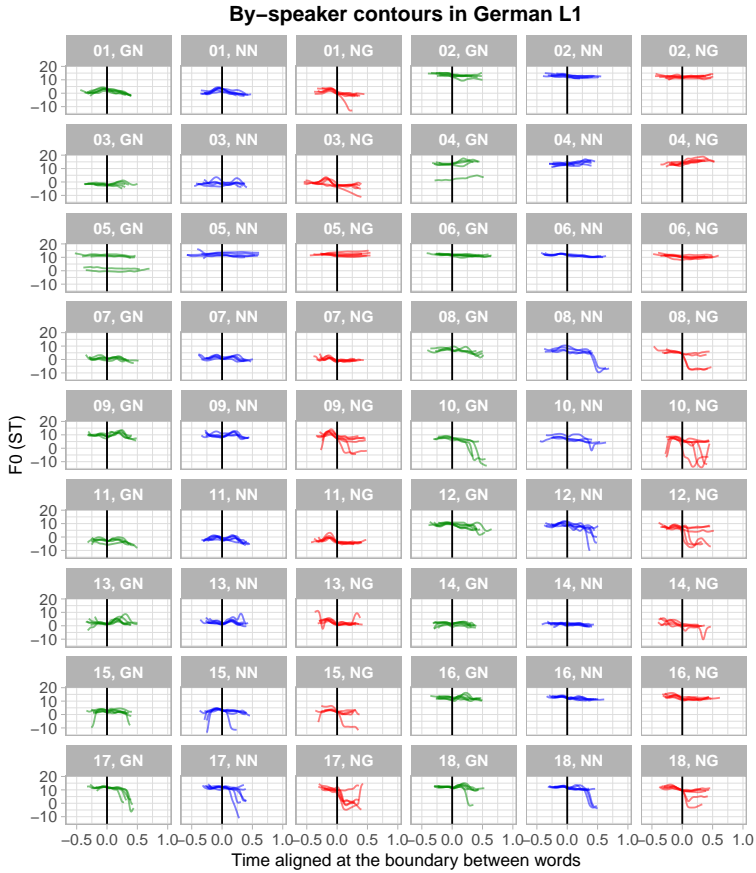


Figure A3: By-speaker F0 contours for the three information structure conditions in L1 German. Information structure conditions are colour-coded: green for given-new (GN), blue for new-new (NN) and red for new-given (NG). The black line marks the boundary between words. Speakers are identified by numbers.

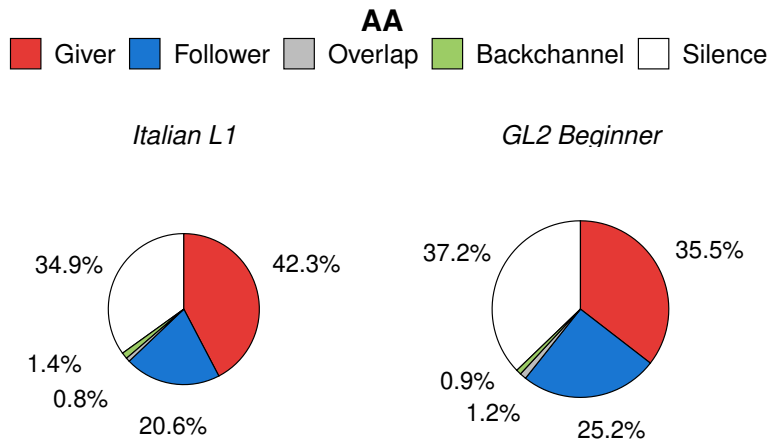


Figure A4: Pie plot summarising conversational activities for dyad AA (L1 Italian and low-proficiency L2 German).

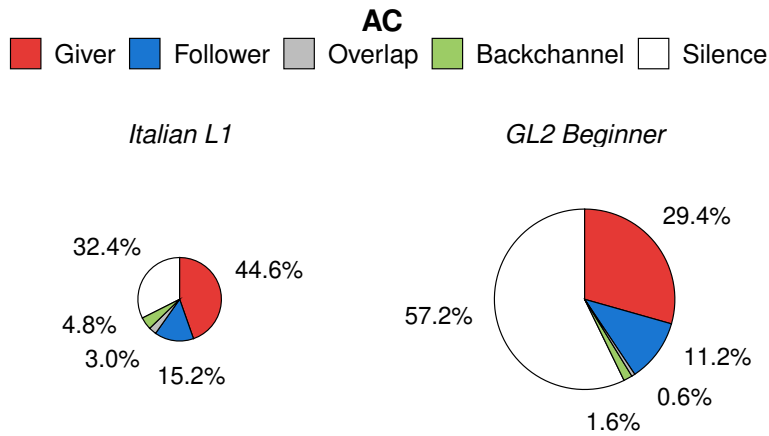


Figure A5: Pie plot summarising conversational activities for dyad AC (L1 Italian and low-proficiency L2 German).

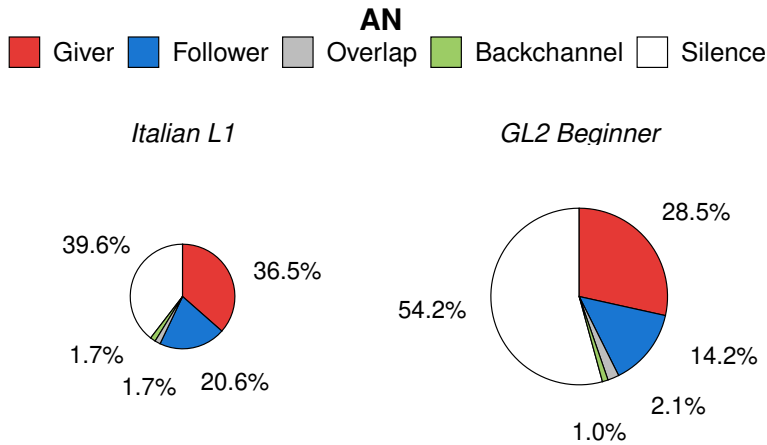


Figure A6: Pie plot summarising conversational activities for dyad AN (L1 Italian and low-proficiency L2 German).

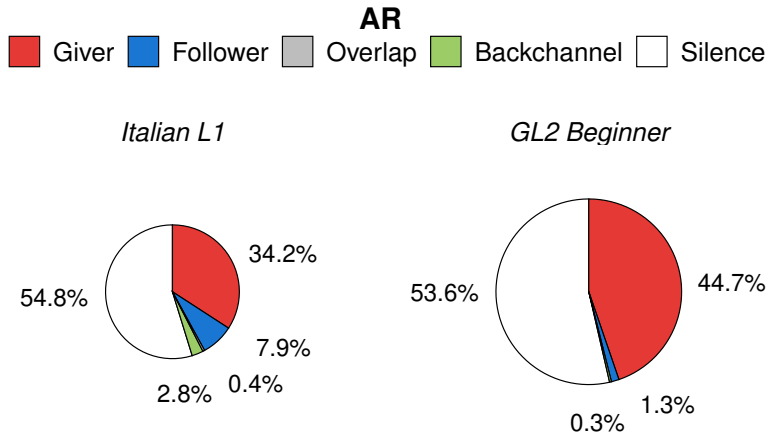


Figure A7: Pie plot summarising conversational activities for dyad AR (L1 Italian and low-proficiency L2 German).

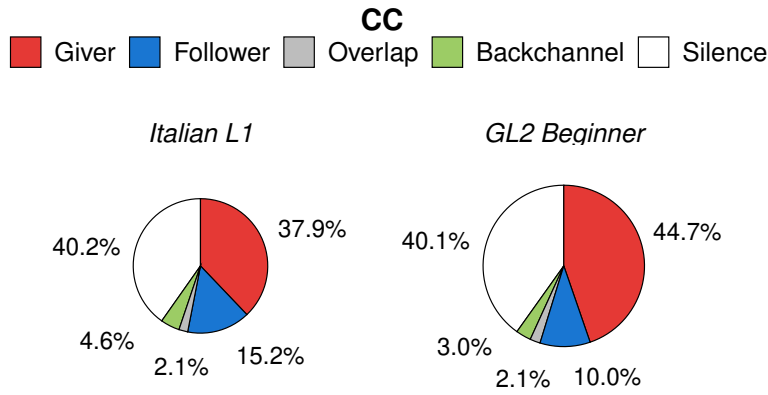


Figure A8: Pie plot summarising conversational activities for dyad CC (L1 Italian and low-proficiency L2 German).

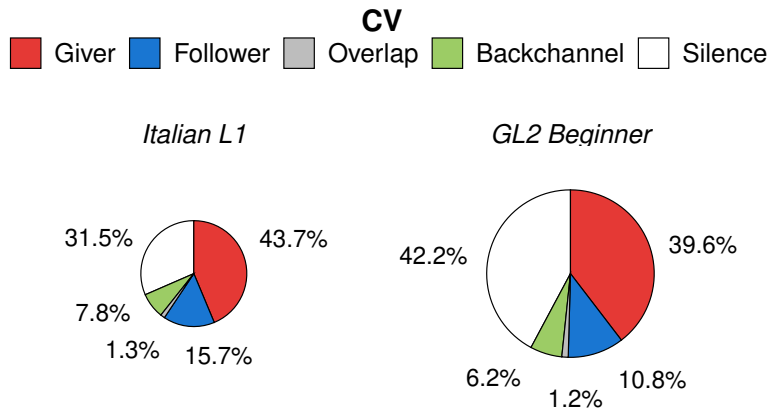


Figure A9: Pie plot summarising conversational activities for dyad CV (L1 Italian and low-proficiency L2 German).

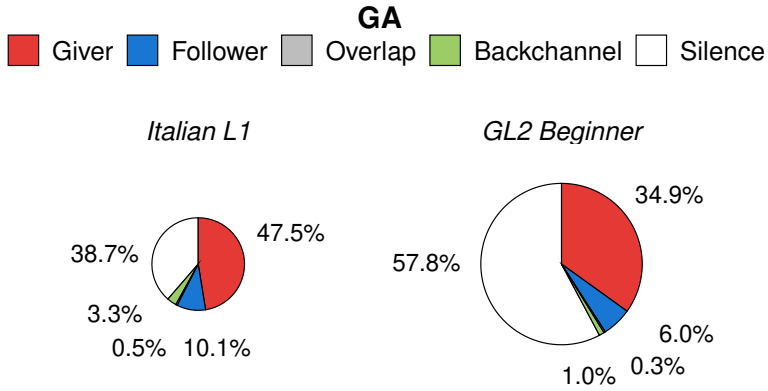


Figure A10: Pie plot summarising conversational activities for dyad GA (L1 Italian and low-proficiency L2 German).

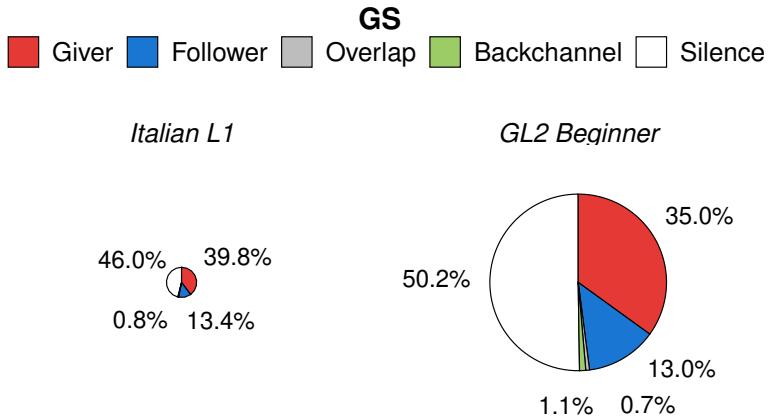


Figure A11: Pie plot summarising conversational activities for dyad GS (L1 Italian and low-proficiency L2 German).

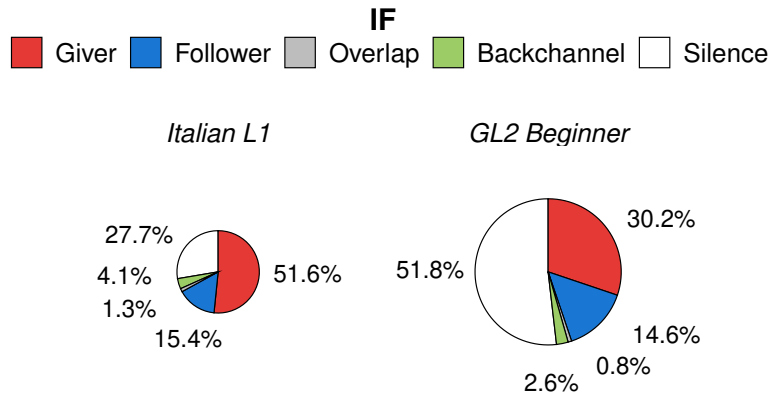


Figure A12: Pie plot summarising conversational activities for dyad IF (L1 Italian and low-proficiency L2 German).

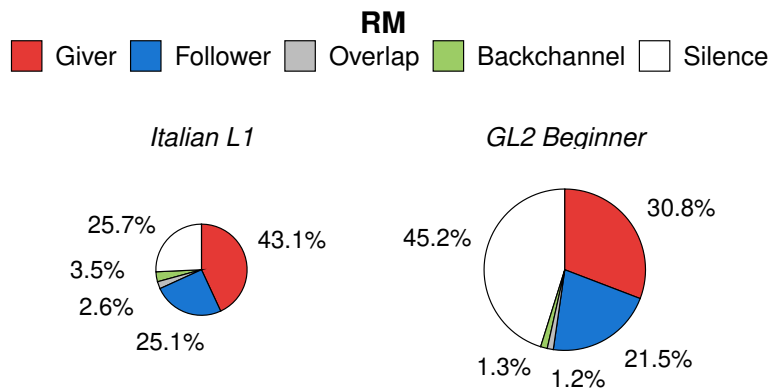


Figure A13: Pie plot summarising conversational activities for dyad RM (L1 Italian and low-proficiency L2 German).

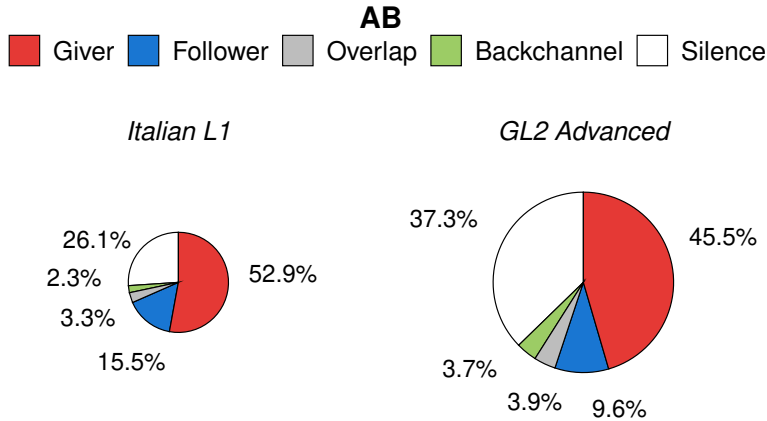


Figure A14: Pie plot summarising conversational activities for dyad AB (L1 Italian and high-proficiency L2 German).

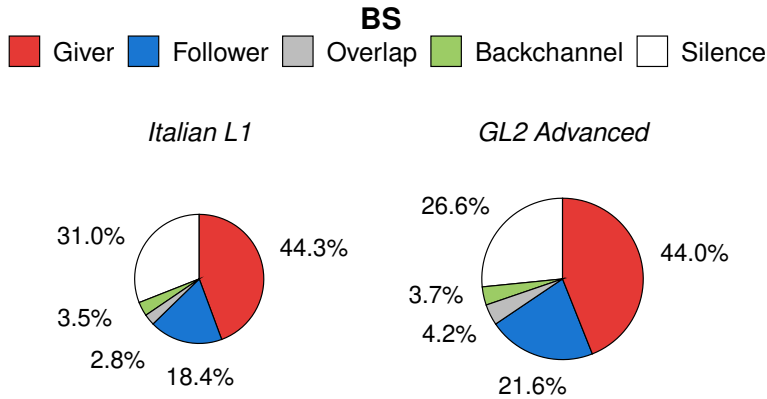


Figure A15: Pie plot summarising conversational activities for dyad BS (L1 Italian and high-proficiency L2 German).

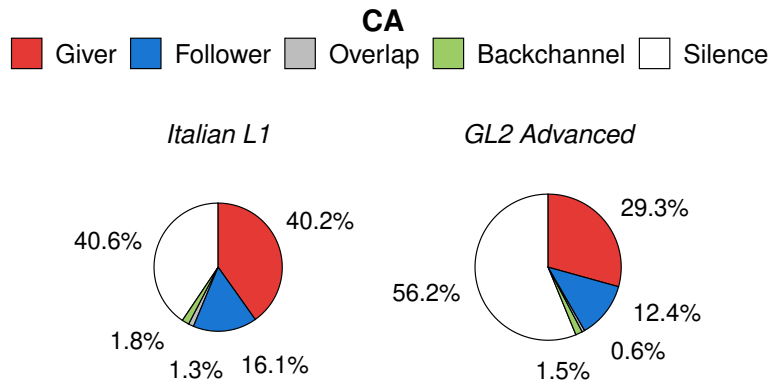


Figure A16: Pie plot summarising conversational activities for dyad CA (L1 Italian and high-proficiency L2 German).

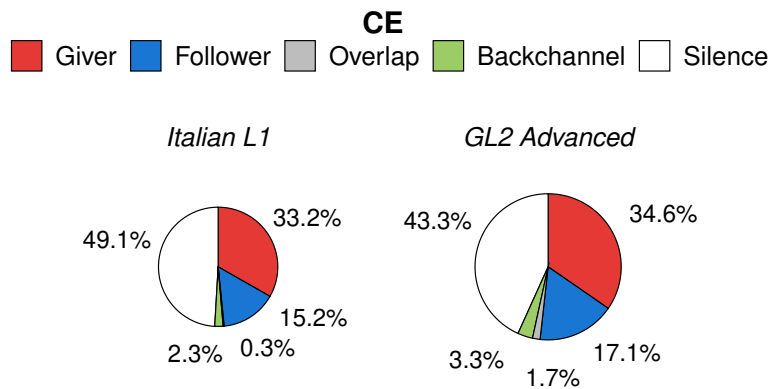


Figure A17: Pie plot summarising conversational activities for dyad CE (L1 Italian and high-proficiency L2 German).

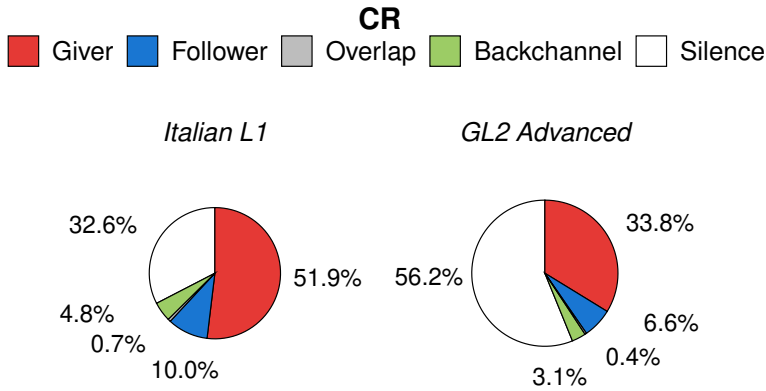


Figure A18: Pie plot summarising conversational activities for dyad CR (L1 Italian and high-proficiency L2 German).

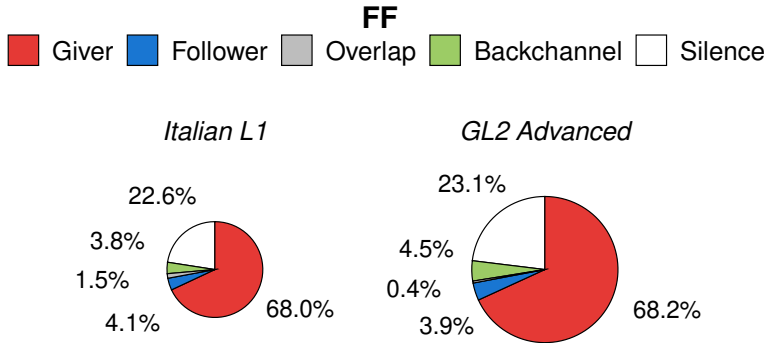


Figure A19: Pie plot summarising conversational activities for dyad FF (L1 Italian and high-proficiency L2 German).

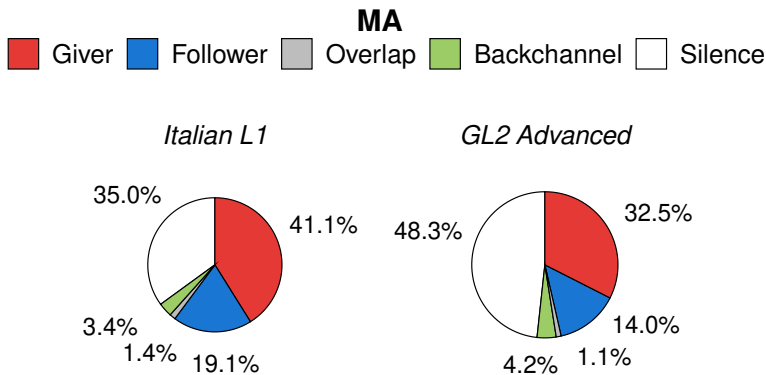


Figure A20: Pie plot summarising conversational activities for dyad MA (L1 Italian and high-proficiency L2 German).

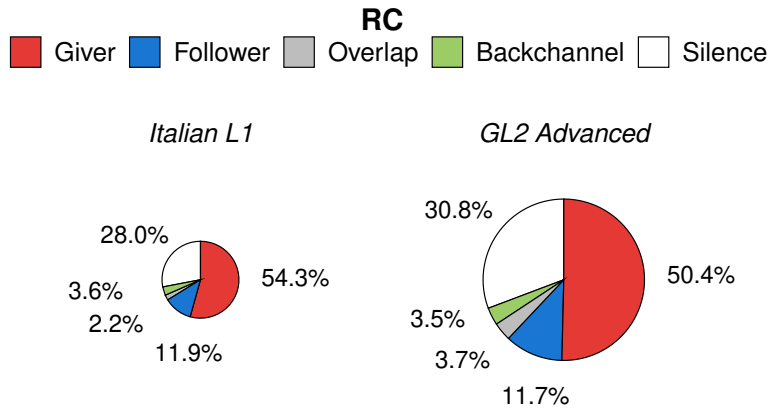


Figure A21: Pie plot summarising conversational activities for dyad RC (L1 Italian and high-proficiency L2 German).

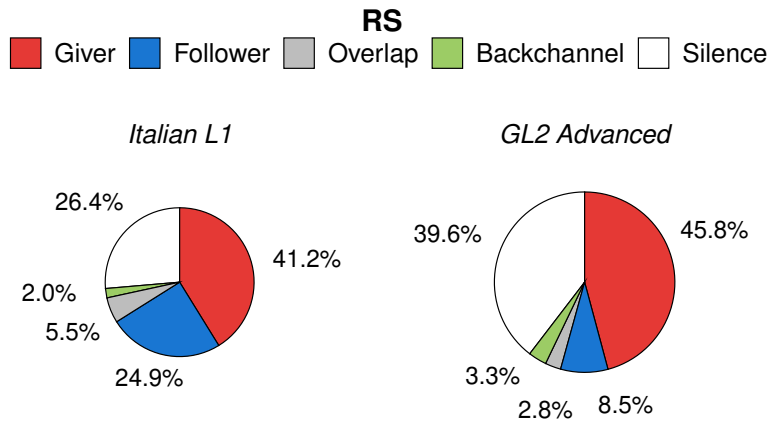


Figure A22: Pie plot summarising conversational activities for dyad RS (L1 Italian and high-proficiency L2 German).

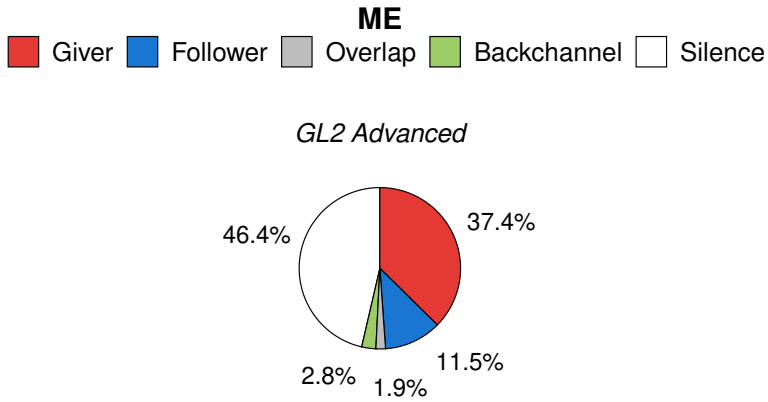


Figure A23: Pie plot summarising conversational activities for dyad ME (L1 Italian and high-proficiency L2 German). Their corresponding file for L2 German was found to be damaged and could not be analysed.

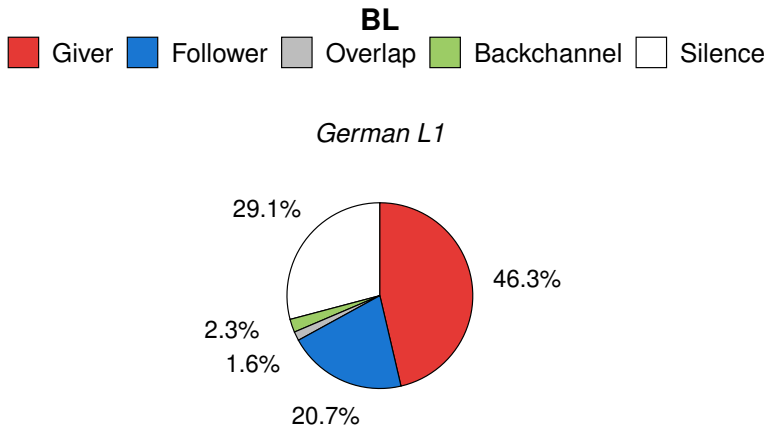


Figure A24: Pie plot summarising conversational activities for dyad BL (L1 German).

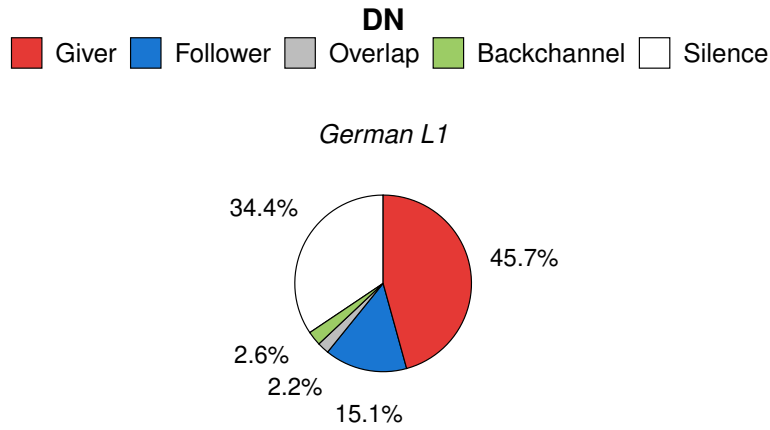


Figure A25: Pie plot summarising conversational activities for dyad DN (L1 German).

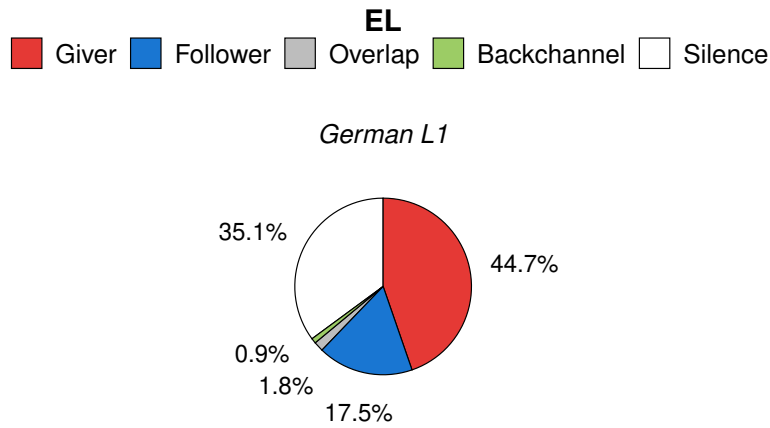


Figure A26: Pie plot summarising conversational activities for dyad EL (L1 German).

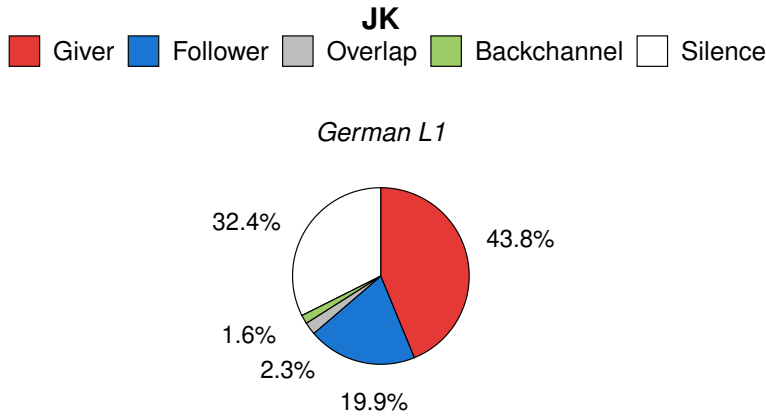


Figure A27: Pie plot summarising conversational activities for dyad JK (L1 German).

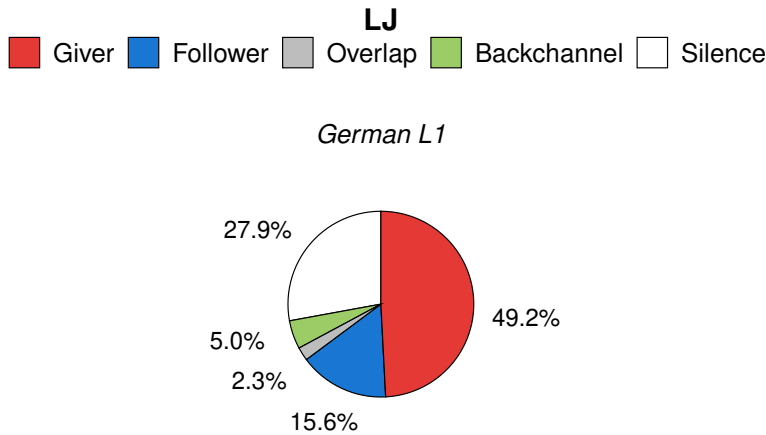


Figure A28: Pie plot summarising conversational activities for dyad LJ (L1 German).

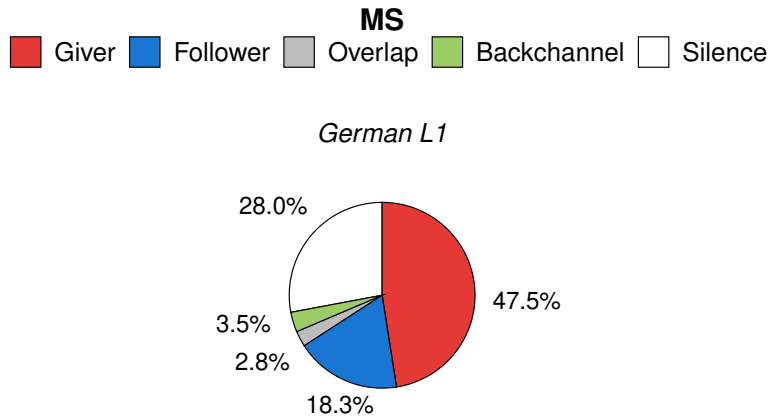


Figure A29: Pie plot summarising conversational activities for dyad MS (L1 German).

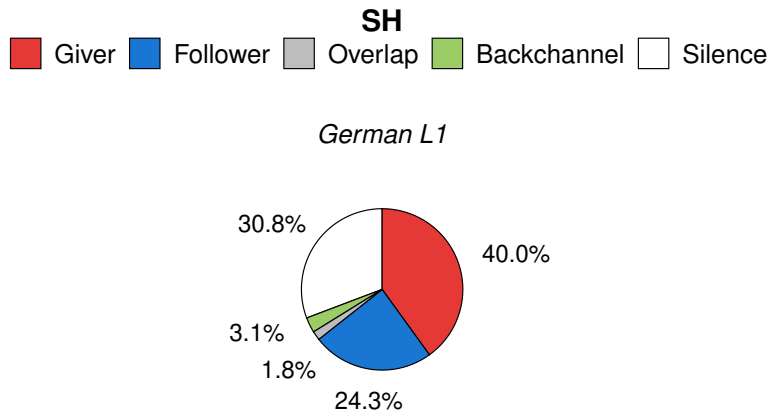


Figure A30: Pie plot summarising conversational activities for dyad SH (L1 German).

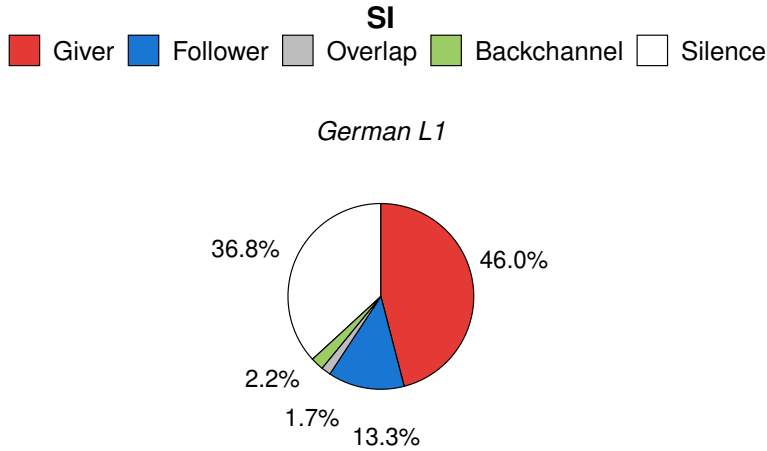


Figure A31: Pie plot summarising conversational activities for dyad SI (L1 German).

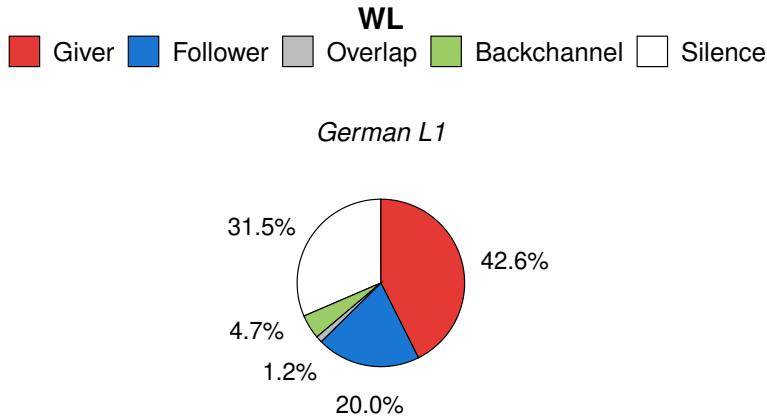


Figure A32: Pie plot summarising conversational activities for dyad WL (L1 German).

References

- Albert, Aviad, Francesco Cangemi, T. Mark Ellison & Martine Grice. 2020. *ProPer: PROsodic analysis with periodic energy [Computer software]*. DOI: 10.17605/OSF.IO/28EA5.
- Albert, Aviad, Francesco Cangemi & Martine Grice. 2018. Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. *Proceedings of Speech Prosody* 9. 13–16.
- Allerton, David J. 1978. The notion of givenness and its relations to presupposition and to theme. *Lingua* 44(1). 133–168.
- Amador-Moreno, Carolina Pilar, Michael McCarthy & Anne O’Keeffe. 2013. Can English provide a framework for Spanish response tokens? In Jesús Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2013*, 175–201. Springer.
- Anderson, Anne, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, et al. 1991. The HCRC map task corpus. *Language and Speech* 34(4). 351–366.
- Aronsson, Berit. 2014. Prosody in the foreign language classroom: Always present, rarely practised? *Journal of Linguistics and Language Teaching* 5(2). 207–224.
- Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen. 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics* 26(1). 3–25.
- Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen. 2006. Tonal association and tonal alignment: Evidence from Greek polar questions and contrastive statements. *Language and Speech* 49(4). 421–450.
- Ashby, Jane. 2016. Why does prosody accompany fluency? Re-conceptualizing the role of phonology in reading. In Asaid Khateb & Irit Bar-Kochva (eds.), *Reading Fluency: Current insights from neurocognitive research and intervention studies* (Literacy Studies 12), 65–89. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-30478-6_5.
- Austin, Gavin, Heejin Chang, Nayoung Kim & Eoin Daly. 2022. Prosodic transfer across constructions and domains in L2 inflectional morphology. *Linguistic Approaches to Bilingualism* 12(5). 657–686. DOI: 10.1075/lab.19076.aus.

References

- Avesani, Cinzia. 1997. I toni della RAI: Un esercizio di lettura intonativa. *Incontri del Centro di studi di grammatica italiana*. 659–727.
- Avesani, Cinzia, Giuliano Bocci & Mario Vayra. 2015. Prosody and information status in Italian and German L2 intonation. In Marina Chini (ed.), *Il parlato in [italiano] L2: aspetti pragmatici e prosodici* (Materiali Linguistici), 93–116. Milano: Franco Angeli.
- Avesani, Cinzia, Giuliano Bocci, Mario Vayra & Alessandra Zappoli. 2013. All'interfaccia tra prosodia e struttura informativa: La realizzazione prosodica dell'informazione data in tedesco e in italiano L2. In Valentina Galatà (ed.), *Multimodalità e multilingualità: la sfida più avanzata della comunicazione orale*, 53–71. Roma: Bulzoni.
- Avesani, Cinzia, Julia Hirschberg & Pilar Prieto. 1995. The intonational disambiguation of potentially ambiguous utterances in English, Italian and Spanish. In *Proceedings of the XIIIth International Congress of Phonetic Sciences*, vol. 1, 174–177.
- Avesani, Cinzia & Mario Vayra. 2005. Accenting, deaccenting and information structure in Italian dialogue. In *6th SIGdial Workshop on Discourse and Dialogue (SIGDial 2005)*, 19–24.
- Bachman, Lyle F. & Adrian S. Palmer. 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Baills, Florence, Charlotte Alazard-Guiu & Pilar Prieto. 2022. Embodied prosodic training helps improve accentedness and suprasegmental accuracy. *Applied Linguistics* 43(4). 776–804. DOI: <https://doi.org/10.1093/applin/amac010>.
- Bard, Ellen Gurman & Matthew Aylett. 1999. The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. In *ICPhS-14*, 1753–1756. San Francisco.
- Baumann, Stefan. 2005. Degrees of givenness and their prosodic marking. In *International Symposium on Discourse and Prosody as a complex interface*.
- Baumann, Stefan. 2006. Information structure and prosody: Linguistic categories for spoken language annotation. *Methods in Empirical Prosody Research* 3. 153–180.
- Baumann, Stefan & Martine Grice. 2006. The intonation of accessibility. *Journal of Pragmatics* 38(10). 1636–1657.
- Baumann, Stefan, Martine Grice & Susanne Steindamm. 2006. Prosodic marking of focus domains: Categorical or gradient. In *Proceedings of speech prosody*, 301–304. Dresden.
- Baumann, Stefan & Kerstin Hadelich. 2003. Accent type and givenness: An experiment with auditory and visual priming. *Proceedings of the 15th ICPhS, Barcelona, Spain*. 1811–1814.

- Baumann, Stefan & Arndt Riester. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. *Prosody And Meaning* 25. 119–162.
- Baumann, Stefan & Arndt Riester. 2013. Coreference, lexical givenness and prosody in German. *Lingua* 136. 16–37.
- Baumann, Stefan & Bodo Winter. 2018. What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics* 70. 20–38.
- Beattie, Geoffrey. 1980. Encoding units in spontaneous speech: Some implications for the dynamics of conversation. In Hans Dechert & Manfred Raupach (eds.), *Temporal variables in speech*, 131–144. De Gruyter Mouton. DOI: 10.1515/9783110816570.131.
- Beckman, Mary & Gayle Ayers. 1997. Guidelines for ToBI labelling. *The OSU Research Foundation* 3(30). 255–309.
- Beckman, Mary, Julia Hirschberg & Stefanie Shattuck-Hufnagel. 2005. The Original ToBI System and the evolution of the ToBI framework. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 9–54. Oxford: Oxford University Press.
- Beckman, Mary & Janet Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology* 3. 255–309.
- Beňuš, Štefan, Agustín Gravano & Julia Hirschberg. 2007. The prosody of backchannels in American English. In Jürgen Trouvain & William J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, 1065–1068. Saarbrücken. DOI: <https://doi.org/10.7916/D8WW7SIP>.
- Berry, Anne. 1994. Spanish and American turn-taking styles: A comparative study. *Pragmatics and Language Learning Monograph Series* (5). 180–190.
- Bertinetto, Pier Marco & Michele Loporcaro. 2005. The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association* 35(2). 131–151.
- Bi, Ran. 2008. English majors' pronunciation and intonation in read speech: A longitudinal study. *Journal of PLA University of Foreign Languages* 31(4). 43–49.
- Bissiri, Maria Paola. 2008. *Experimentelle Untersuchungen zum computergestützten Erlernen der Prosodie des Deutschen für italienischsprachige Lernende*. Ludwig-Maximilians-Universität München. (Doctoral dissertation).
- Bocci, Giuliano. 2013. *The syntax-prosody interface: A cartographic perspective with evidence from Italian* (Linguistik Aktuell / Linguistics Today 204). Amsterdam: John Benjamins Publishing.

References

- Bocci, Giuliano & Cinzia Avesani. 2011. Phrasal prominences do not need pitch movements: Postfocal phrasal heads in Italian. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech)*, 1357–1360. Florence.
- Boersma, Paul & David Weenink. 2021. *Praat: Doing phonetics by computer [Computer Programme]*. Version 6.2.
- Bögels, Sara & Francisco Torreira. 2015. Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics* 52. 46–57.
- Bosker, Hans Rutger, Anne-France Pinget, Hugo Quené, Ted Sanders & Nivja H. De Jong. 2013. What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing* 30(2). 159–175.
- Brazil, David, Malcolm Coulthard & Catherine Johns. 1980. *Discourse intonation and language teaching*. New York: Longman.
- Büring, Daniel. 2007. Semantics, intonation and information structure. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford handbook of linguistic interfaces*, 445–474. Oxford: Oxford University Press.
- Büring, Daniel. 2016. *Intonation and meaning*. Oxford: Oxford University Press.
- Bürkner, Paul-Christian. 2016. *Brms: Bayesian regression models using Stan [R package]*.
- Busà, Maria Grazia. 2008. Teaching prosody to Italian learners of English: Working towards a new approach. In Christoph Taylor (ed.), *Ecolingua: The role of e-corpora in translation and language learning*, 113–126. Trieste: EUT Edizioni Università di Trieste.
- Butterworth, Brian. 1975. Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research* 4. 75–87.
- Calhoun, Sasha. 2009. What makes a word contrastive? Prosodic, semantic and pragmatic perspectives. In Dagmar Barth-Weingarten, Nicole Dehe & Anne Wichmann (eds.), *Where prosody meets pragmatics: Research at the interface*, 53–78. Bingley: Emerald Group Publishing Limited.
- Calhoun, Sasha. 2010. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language* 86. 1–42.
- Campbell, Nick. 1995. Loudness, spectral tilt, and perceived prominence in dialogues. In *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS-1995)*, 676–679. Stockholm.
- Campbell, Nick. 2007. Approaches to conversational speech rhythm: Speech activity in two-person telephone dialogues. In *Proc XVIth International Congress of the Phonetic Sciences*, 343–348. Saarbrücken.

- Campbell-Larsen, John. 2015. Interactional competence in second language acquisition. *Kwansei Gakuin University Humanities Review* 19. 265–287.
- Campbell-Larsen, John. 2022. Interactional competence in L2 learning: Is it taught or activated? *Journal of the Graduate School, Kyoto Women's University*. 45–66.
- Canale, Michael & Merrill Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1(1). 1–47.
- Canepari, Luciano. 1980. *Italiano standard e pronunce regionali*. Padova: CLEUP.
- Cangemi, Francesco, Aviad Albert & Martine Grice. 2019. Modelling intonation: Beyond segments and tonal targets. In *Proceedings of the 19th International Congress of Phonetic Sciences*, 572–576. Melbourne.
- Cangemi, Francesco & Stefan Baumann. 2020. Integrating phonetics and phonology in the study of linguistic prominence. *Journal of Phonetics* 81.
- Cangemi, Francesco, Dina El Zarka, Simon Wehrle, Stefan Baumann & Martine Grice. 2016. Speaker-specific intonational marking of narrow focus in Egyptian Arabic. In *Proceedings of Speech Prosody*, 335–339.
- Cangemi, Francesco & Martine Grice. 2016. The importance of a distributional approach to categoriality in autosegmental-metrical accounts of intonation. *Laboratory Phonology* 7(1). 1–20.
- Cangemi, Francesco, Martine Grice, Alicia Janz, Valeria Lucarini, Malin Spaniol & Kai Vogeley. 2023. Content-free speech activity records: Interviews with people with schizophrenia. *Language Resources and Evaluation*. DOI: 10.1007/s10579-023-09666-z.
- Cangemi, Francesco, Martina Krüger & Martine Grice. 2015. Listener-specific perception of speaker-specific production in intonation. In Susanne Fuchs, Daniel Pape, Caterina Petrone & Pascal Perrier (eds.), *Individual differences in speech production and perception*, 123–145.
- Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li & Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of Statistical Software* 76(1). 1–32.
- Castello, Erik & Sara Gesuato. 2019. Holding up one's end of the conversation in spoken English: Lexical backchannels in L2 examination discourse. *International Journal of Learner Corpus Research* 5(2). 231–252.
- Cekaite, Asta. 2007. A child's development of interactional competence in a Swedish L2 classroom. *The Modern Language Journal* 91. 45–62. DOI: 10.1111/j.1540-4781.2007.00509.x.

References

- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic: a new typology of language*, 25–55. New York: Academic Press.
- Chafe, Wallace. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Chapple, Eliot. 1939. Quantitative analysis of the interaction of individuals. *Proceedings of the National Academy of Sciences* 25(2). 58–67.
- Chodroff, Eleanor, Alaina Arthurs, Priya Kurian, Jonah Pazol & Jennifer Cole. 2019. Categorical and gradient effects of information structure on nuclear prominence in American English. In *Presentation at the 93rd Annual Meeting of the Linguistic Society of America*.
- Chodroff, Eleanor & Jennifer Cole. 2019. The phonological and phonetic encoding of information status in American English nuclear accents. In *Proceedings of the 19th International Congress of Phonetic Sciences*. York.
- Clancy, Patricia, Sandra Thompson, Ryoko Suzuki & Hongyin Tao. 1996. The conversational use of reactive tokens in English, Japanese, and Mandarin. *Journal of Pragmatics* 26(3). 355–387.
- Clark, Herbert & Susan Haviland. 1977. Comprehension and the Given-New Contract. In Roy O. Freedle (ed.), *Discourse processes: Advances in research and theory*, 1–40. New Jersey: Ablex Publishing Corporation.
- Clingwall, Dion, Jon Clenton & Gavin Brooks. 2024. *Exploring relationships that exist between L1 Japanese oral fluency and L2 English oral fluency output*. Rochester. DOI: 10.2139/ssrn.5065870.
- Cohen, Andrew D. 2005. Strategies for learning and performing L2 speech acts. *Intercultural Pragmatics* 2(3). 275–301.
- Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment: Companion volume*. Cambridge: Press Syndicate of the University of Cambridge.
- Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, teaching, assessment: Companion volume*. Strasbourg: Council of Europe Publishing.
- Cruttenden, Alan. 1993. The de-accenting and re-accenting of repeated lexical items. In 16–19. Lund.
- Cruttenden, Alan. 1997. *Intonation*. 2nd edn. Cambridge: Cambridge University Press.
- Cruttenden, Alan. 2006. The de-accenting of given information: A cognitive universal? In Giuliano Bernini & Marcia L. Schwartz (eds.), *Pragmatic organiza-*

- tion of discourse in the languages of europe (Empirical Approaches to Language Typology [EALT] 20), 311–356. Berlin: De Gruyter.
- Cutrone, Pino. 2005. A case study examining backchannels in conversations between Japanese-British dyads. *Multilingua. Journal of Cross-Cultural and Interlanguage Communication* 24(3). 237–274.
- Cutrone, Pino. 2014. A cross-cultural examination of the backchannel behavior of Japanese and Americans: Considerations for Japanese EFL learners. *Intercultural Pragmatics* 11(1). 83–120.
- D’Imperio, Mariapaola. 2000. Acoustic-perceptual correlates of sentence prominence in Italian. *Working Papers in Linguistics* 54. 59–77.
- D’Imperio, Mariapaola. 2002. Italian intonation: An overview and some questions. *Probus* 14. 37–69.
- D’Imperio, Mariapaola & David House. 1997. Perception of questions and statements in Neapolitan Italian. In G. Kokkinakis, N. Fakotakis & E. Dermatas (eds.), *Proceedings of Eurospeech ’97*, 251–254.
- Dahmen, Silvia. 2013. *Prosodie oder Segmente? Phonetische Untersuchungen zu Trainingseffekten bei italienischen Deutschlernenden*. Köln: Universität zu Köln. (Doctoral dissertation).
- De Jong, Nivja H. 2016. Fluency in second language assessment. In Dina Tsagari & Jayanti Banerjee (eds.), *Handbook of second language assessment* (Handbooks of Applied Linguistics [HAL] 12), 203–218. Berlin: De Gruyter Mouton.
- De Jong, Nivja H. 2023. Fluency in speaking as a dynamic construct. *Language Teaching Research Quarterly* 37. 179–187. DOI: 10.32038/ltrq.2023.37.09.
- De Jong, Nivja H., Rachel Groenhout, Rob Schoonen & Jan H. Hulstijn. 2015. Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics* 36(2). 223–243. DOI: 10.1017/S0142716413000210.
- De Jong, Nivja H., Margarita P. Steinel, Arjen Florijn, Rob Schoonen & Jan H. Hulstijn. 2013. Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics* 34(5). 893–916.
- De Ruiter, Laura E. 2015. Information status marking in spontaneous vs. read speech in story-telling tasks: Evidence from intonation analysis using GTotBI. *Journal of Phonetics* 48. 29–44.
- De Saussure, Ferdinand. 1916. Nature of the linguistic sign. In Charles Bally & Albert Sechehaye (eds.), *Cours de linguistique générale*. New York: McGraw Hill Education.
- Derwing, Tracey M. & Murray J. Munro. 2015. *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research* (Language Learning & Language Teaching 42). Amsterdam: John Benjamins.

References

- Derwing, Tracey M., Murray J. Munro & Grace Wiebe. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48(3). 393–410.
- Derwing, Tracey M., Marian J. Rossiter, Murray J. Munro & Ron I. Thomson. 2004. Second language fluency: Judgments on different tasks. *Language Learning* 54(4). 655–679.
- Dingemanse, Mark & Andreas Liesenfeld. 2022. From text to talk: Harnessing conversational corpora for humane and diversity-aware language technology. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, vol. 1: Long Papers, 5614–5633.
- Drummond, Kent & Robert Hopper. 1993. Back channels revisited: Acknowledgment tokens and speakership incipency. *Research on Language and Social Interaction* 26(2). 157–177.
- Duncan, Hilary D., Norman Segalowitz & Natalie A. Phillips. 2014. Differences in L1 linguistic attention control between monolinguals and bilinguals. *Bilingualism: Language and Cognition* 19(1). 106–121. DOI: 10.1017/S136672891400025X.
- Duncan, Starkey. 1974. On the structure of speaker-auditor interaction during speaking turns. *Language in Society* 3(2). 161–180.
- Duncan, Starkey & Donald W. Fiske. 1977. *Face-to-face interaction: Research, methods, and theory*. New Jersey: Lawrence Erlbaum.
- Eckman, Fred R. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27(2). 315–330.
- Eckman, Fred R., Gregory K. Iverson & Jae Yung Song. 2013. The role of hypercorrection in the acquisition of L2 phonemic contrasts. *Second Language Research* 29(3). 257–283.
- Edlund, Jens, Mattias Heldner & Antoine Pelcé. 2010. Prosodic features of very short utterances in dialogue. In *Nordic Prosody-Proceedings of the Xth Conference*, 57–68.
- Ellis, Rod. 1989. Are classroom and naturalistic acquisition the same? A study of the classroom acquisition of German word order rules. *Studies in Second Language Acquisition* 11(3). 305–328.
- Face, Timothy L. & Mariapaola D’Imperio. 2005. Reconsidering a focal typology: Evidence from Spanish and Italian. *Italian Journal of Linguistics/Rivista Di Linguistica* 17(2). 271–289.
- Fang, Wei-Chieh, Hui-Chin Yeh, Bo-Ru Luo & Nian-Shing Chen. 2021. Effects of mobile-supported task-based language teaching on EFL students’ linguistic achievement and conversational interaction. *ReCALL* 33(1). 71–87. DOI: 10.1017/S0958344020000208.

- Farnetani, Edda & Claudio Zmarich. 1997. Prominence patterns in Italian: An analysis of F0 and duration. In Antonis Botinis, Georgios Kouroupetroglou & George Carayannis (eds.), *Intonation: Theory, models and applications. Proceedings of an ESCA Workshop*, 115–118. Athens.
- Felleggy, Anna M. 1995. Patterns and functions of minimal response. *American Speech International Journal of Educational Best Practices* 2(1). 186–199.
- Féry, Caroline. 2013. Focus as prosodic alignment. *Natural Language & Linguistic Theory* 31(3). 683–734.
- Féry, Caroline & Frank Kügler. 2008. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36(4). 680–703.
- Féry, Caroline & Vieri Samek-Lodovici. 2006. Focus projection and prosodic prominence in nested foci. *Language* 82(1). 131–150.
- Figueras, Neus, Brian North, Sauli Takala, Piet Van Avermaet & Norman Verhelst. 2009. Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR): A manual. *Strasbourg: Council of Europe, Language Policy Division*.
- Fillmore, Charles J. 1979. On fluency. In Daniel Kempler & William S.Y. Wang (eds.), *Individual differences in language ability and language behavior*, 85–102. New York: Academic Press.
- Fischer, Kerstin, Oliver Niebuhr & Maria Alm. 2021. Robots for foreign language learning: Speaking style influences student performance. *Frontiers in Robotics and AI* 8.
- Franke, Michael & Timo B. Roettger. 2019. Bayesian regression modeling (for factorial designs): A tutorial. DOI: [10.31234/osf.io/cdxv3](https://doi.org/10.31234/osf.io/cdxv3).
- Fries, Charles C. 1952. *The structure of English*. London: Longmans, Green & Company.
- Frota, Sonia & Pilar Prieto. 2015. *Intonation in Romance*. Oxford: Oxford University Press.
- Galaczi, Evelina D. 2014. Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics* 35(5). 553–574.
- Gardner, Rod. 2001. *When listeners talk: Response tokens and listener stance*. Amsterdam: John Benjamins.
- Gass, Susan M. & Alison J. Mackey. 2014. Input, interaction, and output in second language acquisition. In *Theories in second language acquisition*, 2nd edn. Routledge.
- Gili Fivela, Barbara, Cinzia Avesani, Marco Barone, Giuliano Bocci, Claudia Crocco, Mariapaola D'Imperio, Rosa Giordano, Giovanna Marotta, Michelina Savino & Patrizia Sorianello. 2015. Intonational phonology of the regional va-

References

- ieties of Italian. In Sonia Frota & Pilar Prieto (eds.), *Intonation in Romance*, 140–197. Oxford University Press.
- Gilmartin, Emer, Benjamin R. Cowan, Carl Vogel & Nick Campbell. 2018. Explorations in multiparty casual social talk and its relevance for social human-machine dialogue. *Journal on Multimodal User Interfaces* 12(4). 297–308. DOI: 10.1007/s12193-018-0274-2.
- Giordano, Rosa. 2006. The intonation of polar questions in two central varieties of Italian. In *Proceedings of Speech Prosody*. Dresden.
- Gleitman, Lila R., David January, Rebecca Nappa & John C. Trueswell. 2007. On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language* 57(4). 544–569.
- Goad, Heather & Lydia White. 2019. Prosodic effects on L2 grammars. *Linguistic Approaches to Bilingualism* 9(6). 769–808. DOI: 10.1075/lab.19043.goa.
- Goldman-Eisler, Frieda. 1967. Sequential Temporal Patterns and Cognitive Processes in Speech. *Language and Speech* 10(2). 122–132. DOI: 10.1002/383096701000205.
- Goodwin, Charles. 1986. Audience diversity, participation and interpretation. *Text & Talk* 6. 283–316.
- Gordon, Joshua & Isabelle Darcy. 2022. Teaching segmentals and suprasegmentals: Effects of explicit pronunciation instruction on comprehensibility, fluency, and accentedness. *Journal of Second Language Pronunciation* 8(2). 168–195. DOI: 10.1075/jslp.21042.gor.
- Götz, Sandra. 2013. *Fluency in native and nonnative English speech*. John Benjamins. 1–262.
- Grice, Martine, Stefan Baumann & Ralf Benz Müller. 2005. German intonation in autosegmental-metrical phonology. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 55–83. Oxford: Oxford University Press.
- Grice, Martine, Simon Ritter, Henrik Niemann & Timo B. Roettger. 2017. Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics* 64. 90–107.
- Grice, Martine & Michelina Savino. 2003. Map tasks in Italian: Asking questions about given, accessible and new information. *Catalan Journal of Linguistics* 2. 153–180.
- Grice, Martine, Michelina Savino & Mario Refice. 1997. The intonation of questions in Bari Italian: Do speakers replicate their spontaneous speech when reading? *Phonus* 3(1). 1–7.
- Griffin, Zenzi M. & Kathryn Bock. 2000. What the eyes say about speaking. *Psychological Science* 11(4). 274–279.

- Grosser, Wolfgang. 1997. On the acquisition of tonal and accentual features of English by Austrian learners. In James Allan & Jonathan Leather (eds.), *Second language speech: Structure and process* (Studies on Language Acquisition [SOLA] 13), 211–228. Berlin: De Gruyter.
- Gussenhoven, Carlos. 1983. Focus, mode and the nucleus. *Journal of Linguistics* 19(2). 377–417.
- Gussenhoven, Carlos. 2008. Types of focus in English. In Chungmin Lee, Matthew Gordon & Daniel Büring (eds.), *Topic and focus* (Studies in Linguistics and Philosophy 82), 83–100. Dordrecht: Springer.
- Gut, Ulrike. 2009. *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German* (English Corpus Linguistics 9). Lausanne: Peter Lang.
- Gut, Ulrike & Stefanie Pillai. 2014. Prosodic marking of information structure by Malaysian speakers of English. *Studies in Second Language Acquisition* 36(2). 283–302.
- Gut, Ulrike, Stefanie Pillai & Zuraidah Mohd Don. 2013. The prosodic marking of information status in Malaysian English. *World Englishes* 32(2). 185–197.
- Ha, Kieu-Phuong. 2012. *Prosody in Vietnamese: Intonational form and function of short utterances in conversation*. Canberra: The Australian National University. (Doctoral dissertation).
- Ha, Kieu-Phuong, Samuel Ebner & Martine Grice. 2016. Speech prosody and possible misunderstandings in intercultural talk: A study of listener behaviour in standard Vietnamese and German dialogues. In *Proceedings of Speech Prosody*, 801–805. Boston.
- Hahn, Laura. 2004. Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL quarterly* 38(2). 201–223.
- Hall, Joan Kelly. 1993. The role of oral practices in the accomplishment of our everyday lives: The sociocultural dimension of interaction with implications for the learning of another language. *Applied Linguistics* 14(2). 145–166.
- Hall, Joan Kelly. 1995. (Re)creating our worlds with words: A sociohistorical perspective of face-to-face interaction. *Applied Linguistics* 16(2). 206–232.
- Halliday, Michael. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3(2). 199–244.
- Handley, Zoe Louise. 2024. Exploring the influence of individual speaking style, cross-linguistic transfer of language-specific fluency patterns, and proficiency on second language oral fluency. *Journal of the European Second Language Association* 8(1). 34–47.
- Hasegawa, Yoko. 2014. *Japanese: A linguistic introduction*. Cambridge: Cambridge University Press.

References

- Hawkins, Sarah, Ian Cross & Richard Ogden. 2013. Communicative interaction in spontaneous music and speech. In Martin Orwin, Christine Howes & Ruth Kempson (eds.), *Language, music and interaction* (Communication, Mind and Language 3), 285–329. London: College Publications.
- He, Agnes Weiyun & Richard Young. 1998. Language proficiency interviews: A discourse approach. In Agnes Weiyun He & Richard Young (eds.), *Talking and testing: discourse approaches to the assessment of oral proficiency* (Studies in Bilingualism 14), 1–24. Amsterdam: John Benjamins.
- Heinz, Bettina. 2003. Backchannel responses as strategic responses in bilingual speakers' conversations. *Journal of Pragmatics* 35(7). 1113–1142.
- Heldner, Mattias. 2003. On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics* 31(1). 39–62.
- Henderson, Alan, Frieda Goldman-Eisler & Andrew Skarbek. 1966. Sequential temporal patterns in spontaneous speech. *Language and Speech* 9(4). 207–216. DOI: 10.1177/002383096600900402.
- Heritage, John. 1984. A change-of-state token and aspects of its sequential placement. In J. Maxwell Atkinson (ed.), *Structures of social action: studies in conversation analysis*, 299–345. Cambridge: Cambridge University Press.
- Hermes, Anne, Johannes Becker, Doris Mücke, Stefan Baumann & Martine Grice. 2008. Articulatory gestures and focus marking in German. *Proceedings of Speech Prosody*. 457–460.
- Hermes, Dik J. & H. H. Rump. 1994. Perception of prominence in speech intonation induced by rising and falling pitch movements. *The Journal of the Acoustical Society of America* 96(1). 83–92.
- Hirschberg, Julia & Cinzia Avesani. 1997. The role of prosody in disambiguating potentially ambiguous utterances in English and Italian. In *Intonation: Theory, models and applications: Proceedings of an ESCA Workshop*, 189–192. Athens.
- Höhle, Tilman N. 1992. Über Verum-Fokus im Deutschen. In Joachim Jacobs (ed.), *Informationsstruktur und Grammatik*, 112–141. Wiesbaden: VS Verlag für Sozialwissenschaften. DOI: 10.1007/978-3-663-12176-3_5.
- Huensch, Amanda & Nicole Tracy-Ventura. 2017. L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *The Modern Language Journal* 101(2). 275–293. DOI: 10.1111/modl.12395.
- Hymes, Dell. 1972. On communicative competence. In J.B. Pride & Holmes Janet (eds.), *Sociolinguistics selected reading*, 269–293. London: Penguin Books.
- Ito, Kiwako, Shari Speer & Mary Beckman. 2004. Informational status and pitch accent distribution in spontaneous dialogues in English. In *Proceedings of Speech Prosody*.

- Jacoby, Sally & Elinor Ochs. 1995. Co-construction: An introduction. *Research on Language and Social Interaction* 28. 171–183. DOI: 10.1207/s15327973rlsi2803_1.
- Jefferson, Gail. 1983. Notes on a systematic deployment of the acknowledgement tokens ‘yeah’ and ‘mm hm’. *Tilburg Papers in Language and Literature* 30. 1–18.
- Jilka, Matthias. 2000. *The contribution of intonation to the perception of foreign accent*. Stuttgart: University of Stuttgart. (Doctoral dissertation).
- Jung, Yeon Joo & Scott Crossley. 2024. Stylistic alignment in natural conversation involving second language speakers. *Applied Linguistics Review* 15(3). 871–900. DOI: 10.1515/applirev-2020-0048.
- Jurafsky, Dan, Elizabeth Shriberg, Barbara Fox & Traci Curl. 1998. Lexical, prosodic, and syntactic cues for dialog acts. In *Discourse relations and discourse markers*, 114–120.
- Kahng, Jimin. 2014. Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning* 64(4). 809–854. DOI: 10.1111/lang.12084.
- Kallio, Heini, Maria Kautonen & Mikko Kuronen. 2023. Prosody and fluency of Finland Swedish as a second language: Investigating global parameters for automated speaking assessment. *Speech Communication* 148. 66–80. DOI: 10.1016/j.specom.2023.02.003.
- Kang, Okim. 2010. Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* 38(2). 301–315.
- Kang, Okim, Don Rubin & Lucy Pickering. 2010. Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal* 94(4). 554–566.
- Karttunen, Lauri. 1974. Presupposition and linguistic context. *Theoretical Linguistics* 1(1-3). 181–194.
- Kelly, Niamh. 2022. Change across time in L2 intonation vs. segments: A longitudinal study of the English of Ole Gunnar Solskjaer. *Languages* 7(3).
- Kendon, Adam. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26. 22–63.
- Kim, Yoo Lae, Chen Liu, Pavel Trofimovich & Kim McDonough. 2024. Is nonverbal behavior during conversation related to perceived fluency? *TESOL Journal*.
- Kjellmer, Göran. 2009. Where do we backchannel?: On the use of mm, mhm, uh huh and such like. *International Journal of Corpus Linguistics* 14(1). 81–112.
- Klassen, Jeffrey, Michael Wagner, Annie Tremblay & Heather Goad. 2016. Prominence shifts in English and Spanish parallel constructions. In Julie Hunter, Mandy Simons & Matthew Stone (eds.), *JerSem: The 20th Workshop on the Semantics and Pragmatics of Dialogue (SemDial)*, 76–85.

References

- Kley, Katharina, Silvia Kunitz & Meng Yeh. 2023. L1–L2 speaker interaction: Affordances for assessing repair practices. In M. Rafael Salaberry, Albert Weidemann & Wei-Li Hsu (eds.), *Ethics and context in second language testing*. 186–219. London: Routledge. DOI: 10.4324/9781003384922.
- Kochanski, Greg, Esther Grabe, John Coleman & Burton Rosner. 2005. Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America* 118(2). 1038–1054.
- Kohler, Klaus J. 1991. Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology and semantics. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 25(1). 15–185.
- Kormos, Judit. 2006. *Speech production and second language acquisitions*. Mahwah: Lawrence Erlbaum Associates Publishers.
- Kormos, Judit & Mariann Dénes. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32(2). 145–164.
- Kosmala, Loulou. 2024. *Beyond disfluency: The interplay of speech, gesture, and interaction* (Advances in Interaction Studies 11). Amsterdam: John Benjamins.
- Kousidis, Spyros & David Dorran. 2009. Monitoring convergence of temporal features in spontaneous dialogue speech. In *Dublin Institute of Technology: Digital media Centre Conference Papers*.
- Kraaz, Michelle & Tobias Bernaisch. 2022. Backchannels and the pragmatics of South Asian Englishes. *World Englishes* 41(2). 224–243.
- Krahmer, Emiel & Marc Swerts. 2001. On the alleged existence of contrastive accents. *Speech Communication* 34(4). 391–405.
- Krahmer, Emiel & Marc Swerts. 2008. Perceiving focus. In Chungmin Lee, Matthew Gordon & Daniel Büring (eds.), *Topic and focus: Studies in linguistics and philosophy* (Studies in Linguistics and Philosophy 82), 121–137. Dordrecht: Springer.
- Kramsch, Claire. 1986. From language proficiency to interactional competence. *The Modern Language Journal* 70(4). 366–372.
- Kraut, Robert E., Steven H. Lewis & Lawrence W. Swezey. 1982. Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology* 43(4). 718–731.
- Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3). 243–276.
- Kügler, Frank & Sasha Calhoun. 2020. Prosodic encoding of information structure: A typological perspective. In Carlos Gussenhoven & Aoju Chen (eds.), *The Oxford handbook of language prosody* (Oxford Handbooks), 454–467. Oxford: Oxford University Press.

- Kügler, Frank & Caroline Féry. 2017. Postfocal downstep in German. *Language and Speech* 60(2). 260–288.
- Kuno, Susumu. 1978. Generative discourse analysis in America. In Wolfgang Dressler (ed.), *Current trends in textlinguistics* (Research in Text Theory 2), 275–294. Berlin: De Gruyter.
- Kurumada, Chigusa & Timo B. Roettger. 2022. Thinking probabilistically in the study of intonational speech prosody. *Wiley Interdisciplinary Reviews: Cognitive Science* 13(1).
- Ladd, D. Robert. 1980. *The structure of intonational meaning: Evidence from English*. Bloomington: Indiana University Press.
- Ladd, D. Robert. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
- Ladd, D. Robert. 2008. *Intonational phonology*. 2nd edn. Cambridge: Cambridge University Press.
- Lahmann, Cornelia, Rasmus Steinkrauss & Monika S. Schmid. 2017. Speed, breakdown, and repair: An investigation of fluency in long-term second-language speakers of English. *International Journal of Bilingualism* 21(2). 228–242. DOI: 10.1177/1367006915613162.
- Lambertz, Kathrin. 2011. Back-channelling: The use of *yeah* and *mm* to portray engaged listenership. *Griffith Working Papers in Pragmatics and Intercultural Communication* 4. 11–18.
- Lambrecht, Knud. 1994. *Information structure and sentence form*. Cambridge: Cambridge University Press.
- Leiner, Dominik J. 2019. *SoSci Survey [Computer software]*. Version 3.1.06.
- Lemhöfer, Kristin & Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods* 44(2). 325–343. <http://www.lextale.com/>.
- Lemoine, Nathan P. 2019. Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos* 128(7). 912–928.
- Lengeris, Angelos. 2012. Prosody and second language teaching: Lessons from L2 speech perception and production research. In Jesús Romero-Trillo (ed.), *Pragmatics and prosody in English language teaching* (Educational Linguistics 15), 25–40. Berlin: Springer.
- Lennon, Paul. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 40(3). 387–417.
- Lennon, Paul. 2000. The lexical element in spoken second language fluency. In Heidi Riggensbach (ed.), *Perspectives on fluency*, 25–42. Ann Arbor: The University of Michigan Press.

References

- Lepschy, Anna Laura & Guilio Lepschy. 1977. *The Italian language today*. London: Hutchinson.
- LeVelle, Kimberly & John Levis. 2014. Understanding the impact of social factors on L2 pronunciation: Insights from learners. In John M. Levis & Alene Moyer (eds.), *Social dynamics in second language accent*, 97–118. Berlin: De Gruyter. DOI: 10.1515/9781614511762.97.
- Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, Stephen C. 2016. Turn-taking in human communication: Origins and implications for language processing. *Trends in cognitive sciences* 20(1). 6–14.
- Levinson, Stephen C. & Francisco Torreira. 2015. Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology* 6. DOI: 10.3389/fpsyg.2015.00731.
- Lewis, David & Arnim Stechow. 1979. Scorekeeping in a language game. In Rainer Bäuerle & Urs Egli (eds.), *Semantics from different points of view* (Springer Series in Language and Communication 6), 172–187. Berlin: Springer.
- Li, Han Z. 2006. Backchannel responses as misleading feedback in intercultural discourse. *Journal of Intercultural Communication Research* 35(2). 99–116.
- Li, Han Z., Yanping Cui & Zhizhang Wang. 2010. Backchannel responses and enjoyment of the conversation: The more does not necessarily mean the better. *International Journal of Psychological Studies* 2(1). 25–37.
- Loewen, Shawn & Masatoshi Sato. 2018. Interaction and instructed second language acquisition. *Language teaching* 51(3). 285–329.
- Logan, Kenneth J. 2015. *Fluency disorders*. San Diego: Plural Publishing.
- Long, Mike. 1985. A role for instruction in second language acquisition: Task-based language teaching. In Kenneth Hyldenstam & Manfred Pienemann (eds.), *Modelling and assessing second language acquisition*, vol. 18 (Multilingual Matters), 77–99. Clevedon: Multilingual Matters.
- Long, Mike. 2015. Tblt: Building the road as we travel. In Martin Bygate (ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference*, 1–26. John Benjamins. DOI: 10.1075/tblt.8.01lon.
- Macdonald, Shem. 2002. Pronunciation-views and practices of reluctant teachers. *Prospect* 17(3). 3–18.
- Mackey, Alison J., Nicole Ziegler & Lara Bryfonski. 2016. From SLA research on interaction to TBLT materials. In Brian Tomlinson (ed.), *SLA Research and Materials Development for Language Learning*. London: Taylor & Francis.
- Magne, Viktoria, Shungo Suzuki, Yui Suzukida, Meltem Ilkan, Mai Tran & Kazuya Saito. 2019. Exploring the dynamic nature of second language listeners' perceived fluency: A mixed-methods approach. *TESOL Quarterly* 53(4). 1139–1150.

- Magno-Caldognetto, Emanuela, Franco E. Ferrero, Carlo Lavagnoli & Kyriaki Vaggas. 1978. F0 contours of statements, yes-no questions, and wh-questions of two regional varieties of Italian. *Journal of Italian Linguistics Amsterdam* 3(1). 57–66.
- Major, Roy C. & Eunyi Kim. 1996. The similarity differential rate hypothesis. *Language Learning* 46(3). 465–496.
- May, Lyn, Fumiyo Nakatsuhara, Daniel Lam & Evelina Galaczi. 2020. Developing tools for learning-oriented assessment of interactional competence: bridging theory and practice. *Language Testing* 37(2). 165–188. DOI: 10.1177/0265532219879044.
- McCarthy, Michael. 2009. Rethinking spoken fluency. *Estudios de lingüística inglesa aplicada* 9. 11–29.
- McDonough, Kim, Rachael Lindberg, Yoo Lae Kim & Pavel Trofimovich. 2024. Raising business communication students' awareness of nonverbal features of interaction. *Language Awareness* 33(2). 224–239. DOI: 10.1080/09658416.2023.2195666.
- McDonough, Kim, Pavel Trofimovich, Libing Lu & Dato Abashidze. 2020. Visual cues during interaction: Are recasts different from noncorrective repetition? *Second Language Research* 36(3). 359–370. DOI: 10.1177/0267658320914962.
- Mennen, Ineke. 1998. Can language learners ever acquire the intonation of a second language? *STiLL: Speech Technology in Language Learning* 98. 17–20.
- Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32(4). 543–563.
- Missaglia, Federica. 2007. Prosodic training for adult Italian learners of German: The contrastive prosody method. In Jürgen Trouvain & Ulrike Gut (eds.), *Non-native prosody: Phonetic description and teaching practice* (Trends in Linguistics: Studies and Monographs [TiLSM] 186), 236–258. Berlin: De Gruyter.
- Morrison, Astrid. 2018. *Fluency in the EFL Chilean classrooms: To what extent do teachers and textbooks promote oral fluency?* Master's dissertation, University of Reading. London: British Council.
- Mücke, Doris & Martine Grice. 2014. The effect of focus marking on supralaryngeal articulation: Is it mediated by accentuation? *Journal of Phonetics* 44. 47–61.
- Munro, Murray J. 1995. Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition* 17(1). 17–34.
- Munro, Murray J. 2003. A primer on accent discrimination in the Canadian context. *TESL Canada Journal* 20(2). 38–51.

References

- Munro, Murray J. & Tracey M. Derwing. 1999. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45(1). 73–97.
- Munro, Murray J. & Tracey M. Derwing. 2001. Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Studies in Second Language Acquisition* 23(4). 451–468.
- Munro, Murray J. & Tracey M. Derwing. 2020. Foreign accent, comprehensibility and intelligibility, redux. *Journal of Second Language Pronunciation* 6(3). 283–309. DOI: 10.1075/jslp.20038.mun.
- Nakane, Ikuko. 2007. *Silence in intercultural communication*. Amsterdam: John Benjamins Publishing.
- Nava, Emily A. 2010. *Connecting phrasal and rhythmic events: Evidence from second language speech*. Los Angeles: University of Southern California.
- Nava, Emily A. & Maria Luisa Zubizarreta. 2008. Prosodic transfer in L2 speech: Evidence from phrasal prominence and rhythm. *Proceedings of Speech Prosody 2008*. 335–338.
- Nicora, Francesca, Laura Incalcaterra McLoughlin & Barbara Gili Fivela. 2018. Impact of prosodic training on Italian as L2 by Hiberno-English speakers: The case of polar questions. In *Proceedings of Speech Prosody 2018*.
- Niebuhr, Oliver, Maria Alm, Nathalie Schümchen & Kerstin Fischer. 2017. Comparing visualization techniques for learning second language prosody: First results. *International Journal of Learner Corpus Research* 3(2). 250–277.
- Niebuhr, Oliver, Mariapaola D’Imperio, Barbara Gili Fivela & Francesco Cangemi. 2011. Are there ‘shapers’ and ‘aligners’? Individual differences in signalling pitch accent category. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 120–123. Hong Kong.
- Nurjaleka, Lisda. 2019. Backchannel behavior in interview discourse: A contrastive study between Japanese and Indonesian. In *Eleventh Conference on Applied Linguistics (CONAPLIN 2018)*, 451–457. Atlantis Press.
- Paul, Hermann. 2013. *Mittelhochdeutsche Grammatik*. Berlin: De Gruyter.
- Pekarek Doehler, Simona & Evelyne Berger. 2015. The development of L2 interactional competence: Evidence from turn-taking organization, sequence organization, repair organization and preference organization. In Teresa Cadierno & Søren Wind Eskildsen (eds.), *Usage-based perspectives on second language learning*, 233–268. De Gruyter Mouton. DOI: 10.1515/9783110378528-012.
- Pekarek Doehler, Simona & Evelyne Berger. 2018. L2 interactional competence as increased ability for context-sensitive conduct: A longitudinal study of story-openings. *Applied Linguistics* 39(4). 555–578. DOI: 10.1093/applin/amw021.

- Peltonen, Pauliina. 2017. L2 fluency in spoken interaction: A case study on the use of other-repetitions and collaborative completions. *AFinLA-teema* (10). 118–138. DOI: 10.30660/afinla.73130.
- Peltonen, Pauliina. 2020. *Individual and interactional speech fluency in L2 English from a problem-solving perspective: A mixed-methods approach*. Turku. (Doctoral dissertation).
- Peltonen, Pauliina. 2022. Connections between measured and assessed fluency in L2 peer interaction: A problem-solving perspective. en. *International Review of Applied Linguistics in Language Teaching* 60(4). 983–1011. DOI: 10.1515/iral-2020-0030.
- Peltonen, Pauliina. 2024. Fluency revisited. *ELT Journal* 78(4). 489–492. DOI: 10.1093/elt/ccad047.
- Peltonen, Pauliina & Pekka Lintunen. 2016. Integrating quantitative and qualitative approaches in L2 fluency analysis: A study of Finnish-speaking and Swedish-speaking learners of English at two school levels. *European Journal of Applied Linguistics* 4(2). 209–238. DOI: 10.1515/eujal-2014-0018.
- Pérez, Nubia Patricia Carrero. 2016. Effects of tasks on spoken interaction and motivation in English language learners. *GIST – Education and Learning Research Journal* (13). 34–55. DOI: 10.26817/16925777.311.
- Petrov, Sergei V. 2021. Hypercorrection in Russian as a foreign language. In Ekaterina Bakshutova, Viktoriya Dobrova & Yulia Lopukhova (eds.), *Humanity in the era of uncertainty: European proceedings of social and behavioural sciences*, vol. 119, 344–350. European Publisher. DOI: 10.15405/epsbs.2021.12.02.43.
- Piccardo, Enrica, Tom Goodier & Brian North. 2018. *Common European Framework of Reference for Languages: Learning, teaching, assessment: Companion volume with new descriptors*. Strasbourg: Council of Europe Publishing.
- Piccardo, Enrica & Brian North. 2017. Developing phonology descriptors for the Common European Framework of Reference (CEFR). *Pronunciation in Second Language Learning and Teaching Proceedings* 8(1). 97–109.
- Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*. Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).
- Pika, Simone, Ray Wilkinson, Kobin H. Kendrick & Sonja C. Vernes. 2018. Taking turns: Bridging the gap between human and animal communication. *Proceedings of the Royal Society B: Biological Sciences* 285(1880). DOI: 10.1098/rspb.2018.0598.
- Piske, Thorsten. 2012. Factors affecting the perception and production of L2 prosody: Research results and their implications for the teaching of foreign languages. In Jesús Romero-Trillo (ed.), *Pragmatics and prosody in English lan-*

References

- guage teaching* (Educational Linguistics 15), 41–59. Dordrecht: Springer Netherlands. DOI: 10.1007/978-94-007-3883-6_4.
- Préfontaine, Yvonne, Judit Kormos & Daniel Ezra Johnson. 2016. How do utterance measures predict raters' perceptions of fluency in French as a second language? *Language Testing* 33(1). 53–73.
- Prince, Ellen F. 1981. Towards a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–256. New York: Academic Press.
- Purcell, Edward T. & Richard W. Suter. 1980. Predictors of pronunciation accuracy: A reexamination. *Language Learning* 30(2). 271–287.
- R Core Team. 2013. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- R Core Team. 2021. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rasier, Laurent, Johanneke Caspers & Vincent J. Van Heuven. 2010. Accentual marking of information status in Dutch and French as foreign languages. Production and perception. In Katarzyna Dziubalska-Kolaczyk, Magdalena Wrembel & Malgorzata Kul (eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010*, 379–385.
- Rasier, Laurent & Philippe Hiligsmann. 2007. Prosodic transfer from L1 to L2: Theoretical and methodological issues. *Nouveaux cahiers de linguistique française* 28. 41–66.
- Ravignani, Andrea, Laura Verga & Michael D. Greenfield. 2019. Interactive rhythms across species: The evolutionary biology of animal chorusing and turn-taking. *Annals of the New York Academy of Sciences* 1453(1). 12–21. DOI: 10.1111/nyas.14230.
- Reinhardt, Susanne. 2022. *Assessing interactional competence: Identifying candidate criterial features for L2 repair skills*. Potsdam: Universität Potsdam. (Doctoral dissertation).
- Riester, Arndt & Jörn Piontek. 2015. Anarchy in the NP: When new nouns get deaccented and given nouns don't. *Lingua* 165(B). 230–253.
- Riggenbach, Heidi. 1991. Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes* 14(4). 423–441.
- Ritter, Simon & Martine Grice. 2015. The role of tonal onglides in German nuclear pitch accents. *Language and Speech* 58(1). 114–128.
- Rivers, Wilga M. 1994. Comprehension and production: The interactive duo. In Ronald M. Barasch & C. Vaughn James (eds.), *Beyond the Monitor Model: Comments on current theory and practice in Second Language Acquisition*, 71–95. Boston: Heinle & Heinle.

- Roberts, Benjamin & Kim Kirsner. 2000. Temporal cycles in speech production. *Language and Cognitive Processes* 15(2). 129–157. DOI: 10.1080/016909600386075.
- Roberts, Felicia, Piera Margutti & Shoji Takano. 2011. Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese. *Discourse Processes* 48(5). 331–354.
- Roessig, Simon, Doris Mücke & Martine Grice. 2019. The dynamics of intonation: Categorical and continuous variation in an attractor-based model. *PloS One* 14(5). DOI: 10.1371/journal.pone.0216859.
- Röhr, Christine Tanja & Stefan Baumann. 2010. Prosodic marking of information status in German. In *Proceedings of Speech Prosody*, 1–4. Chicago.
- Rossiter, Marian J. 2009. Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review* 65(3). 395–412.
- Rühlemann, Christoph. 2007. *Conversation in context: A corpus-driven approach*. London: Longman.
- Rühlemann, Christoph & Alexander Ptak. 2023. Reaching beneath the tip of the iceberg: A guide to the Freiburg Multimodal Interaction Corpus. *Open Linguistics* 9(1). DOI: 10.1515/opli-2022-0245.
- Sacks, Harvey, Emanuel A. Schegloff & Gail Jefferson. 1974. A simplest systematics for the organization of turn taking for conversation. *Language* 50(4). 696–735.
- Saito, Kazuya, Meltem Ilkan, Viktoria Magne, Mai Ngoc Tran & Shungo Suzuki. 2018. Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics* 39(3). 593–617.
- Sato, Masatoshi. 2014. Exploring the construct of interactional oral fluency: Second language acquisition and language testing approaches. *System* 45. 79–91. DOI: 10.1016/j.system.2014.05.004.
- Savignon, Sandra J. 1991. Communicative language teaching: State of the art. *TESOL quarterly* 25(2). 261–278.
- Savino, Michelina. 2010. Intonational strategies for backchanneling in Italian Map Task dialogues. In *ExLing*, 157–160.
- Savino, Michelina. 2011. The intonation of backchannels in Italian task-oriented dialogues: Cues to turn-taking dynamics, information status and speakers' attitude. In *Proceedings of the 5th Language and Technology Conference: Human language technology as a challenge for computer science and linguistics*, 370–374. Poznan.
- Savino, Michelina. 2012. The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association* 42(1). 23–48.

References

- Savino, Michelina. 2014. The intonation of backchannel tokens in Italian collaborative dialogues. In Zygmunt Vetulani, Joseph Mariani & Marek Kubis (eds.), *Human language technology challenges for computer science and linguistics*. Berlin: Springer.
- Savino, Michelina & Mario Refice. 2013. Acknowledgement or reply? Prosodic features for disambiguating pragmatic functions of the Italian token 'sì'. In *7th Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, 1–6. IEEE.
- Sbranna, Simona, Aviad Albert & Martine Grice. 2025. Investigating interlanguages beyond categorical analyses: Prosodic marking of information status in Italian learners of German. *Journal of Phonetics* 108.
- Sbranna, Simona, Francesco Cangemi & Martine Grice. 2020. Quantifying L2 interactional competence. In Luciano Romito (ed.), *Language variation under contact conditions: Acquisitional contexts, languages, dialects and minorities in Italy and around the world*, 383–405. XVI AISV Conference Proceedings. Milano: Officinaventuno.
- Sbranna, Simona, Eduardo Möking, Simon Wehrle & Martine Grice. 2022. Backchannelling across Languages: Rate, lexical choice and intonation in L1 Italian, L1 German and L2 German. *Proceedings of Speech Prosody*. 734–738.
- Sbranna, Simona, Michelina Savino, Florence Bails & Martine Grice. 2025. Gaze behaviour and vocal feedback in task-based dyadic conversations with and without eye contact. *Frontiers in Communication* 10. DOI: 10.3389/fcomm.2025.1524453.
- Sbranna, Simona, Caterina Ventura, Aviad Albert & Martine Grice. 2023. Prosodic marking of information status in Italian. *Journal of Phonetics* 97. DOI: 10.1016/j.wocn.2023.101212.
- Sbranna, Simona, Simon Wehrle & Martine Grice. 2024. A multi-dimensional analysis of backchannels in L1 German, L1 Italian and L2 German. *Language, Interaction, and Acquisition* 15(2). 242–276.
- Schegloff, Emanuel A. 1982. Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In Deborah Tannen (ed.), vol. 71, 71–93. Washington, DC: Georgetown University Press.
- Schegloff, Emanuel A. 2006. Interaction: The infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted. In *Roots of Human Sociality*. Routledge.
- Schiffer, Stephen R. 1972. *Meaning*. Oxford: Clarendon Press.
- Schnur, Tatiana T., Albert Costa & Alfonso Caramazza. 2006. Planning at the phonological level during sentence production. *Journal of Psycholinguistic Research* 35(2). 189–213.

- Schweitzer, Katrin, Arndt Riester, Michael Walsh & Grzegorz Dogil. 2009. Pitch accents and information status in a German radio news corpus. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech)*, 2416–2418. Brighton.
- Segalowitz, Norman. 2010. *Cognitive bases of second language fluency*. London: Routledge.
- Segalowitz, Norman. 2016. Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching* 54(2). 79–95.
- Segalowitz, Norman & Barbara F. Freed. 2004. Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies In Second Language Acquisition* 26(2). 173–199.
- Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics in Language Teaching* 10(1-4). 209–232. DOI: 10.1515/iral.1972.10.1-4.209.
- Selkirk, Elisabeth. 1995. Sentence prosody: Intonation, stress, and phrasing. 1. 550–569.
- Senk, K. Maynard. 1997. Analyzing interactional management in native/non-native English conversation: A case of listener response. *IRAL: International Review of Applied Linguistics in Language Teaching* 35(1). 37.
- Shelley, Leah & Fernando Gonzalez. 2013. Back channeling: Function of back channeling and L1 effects on back channeling in L2. *Linguistic Portfolios* 2(1). 98–108.
- Shi, Shuju, Jinsong Zhang & Yanlu Xie. 2014. Cross-language comparison of F0 range in speakers of native Chinese, native Japanese and Chinese L2 of Japanese: Preliminary results of a corpus-based analysis. In *The 9th International Symposium on Chinese Spoken Language Processing*, 241–244. IEEE.
- Shibata, Miki. 2023. Japanese L2 English learners' positions in miscommunication: Who is responsible for failures? *Journal of Language, Identity & Education* 22(6). 589–605. DOI: 10.1080/15348458.2021.1938572.
- Simon, Carsta. 2018. The functions of active listening responses. *Behavioural Processes* 157. 47–53.
- Sluijter, Agaath M. C. & Vincent J. Van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100(4). 2471–2485.
- Smorenburg, Laura, Joe Rodd & Aoju Chen. 2015. The effect of explicit training on the prosodic production of L2 sarcasm by Dutch learners of English. In *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–47. Glasgow.

References

- Song, Jieun & Paul Iverson. 2018. Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition* 179. 163–170. DOI: 10.1016/j.cognition.2018.06.001.
- Sørensen, A. Josefine Munch, Michal Fereczkowski & Ewen MacDonald. 2019. Effects of noise and L2 on the timing of turn taking in conversation. *Proceedings of the International Symposium on Auditory and Audiological Research* 7. 85–92.
- Sørensen, A. Josefine Munch, Michal Fereczkowski & Ewen N. MacDonald. 2021. Effects of noise and second language on conversational dynamics in task dialogue. *Trends in Hearing* 25. DOI: 10.1177/23312165211024482.
- Spaniol, Malin, Alicia Janz, Simon Wehrle, Kai Vogeley & Martine Grice. 2023. Multimodal signalling: The interplay of oral and visual feedback in conversation. In *Proceedings of the 20th International Congress of Phonetic Sciences. ICPhS*.
- Stalnaker, Robert. 1974. Pragmatic presuppositions. In Milton Munitz & Peter Unger (eds.), *Semantics and philosophy*, 197–213. New York: New York University Press.
- Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy* 25(5/6). 701–721.
- Starkweather, C. Woodruff. 1987. *Fluency and stuttering*. Hoboken: Prentice-Hall, Inc.
- Stivers, Tanya, Nicholas J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan Peter De Ruiter, Kyung-Eun Yoon, et al. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106(26). 10587–10592.
- Stocksmeier, Thorsten, Stefan Kopp & Dafydd Gibbon. 2007. Synthesis of prosodic attitudinal variants in German backchannel ja. In *Interspeech*, 1290–1293.
- Suter, Richard W. 1976. Predictors of pronunciation accuracy in second language learning. *Language Learning* 26(2). 233–253.
- Suzuki, Shungo & Judit Kormos. 2020. Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition* 42(1). 143–167.
- Swerts, Marc, Cinzia Avesani & Emiel Krahmer. 1999. Reaccentuation or deaccentuation: a comparative study of Italian and Dutch. In *Proceedings of the 14th ICPhS, San Francisco*, 1541–144.

- Swerts, Marc, Emiel Krahmer & Cinzia Avesani. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30(4). 629–654.
- Swerts, Marc & Sabine Zerbian. 2010. Intonational differences between L1 and L2 English in South Africa. *Phonetica* 67(3). 127–146.
- Takahashi, Daniel Y., Alicia R. Fenley & Asif A. Ghazanfar. 2016. Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371(1693). 20150370. DOI: 10.1098/rstb.2015.0370.
- Takeda, Aya. 2018. *English prosodic marking of information structure by L1 Japanese second language learners*. University of Hawai'i at Manoa. (Doctoral dissertation).
- Tao, Hongyin & Sandra A. Thompson. 1991. English backchannels in Mandarin conversations: A case study of superstratum pragmatic 'interference'. *Journal of Pragmatics* 16(3). 209–223.
- Tavakoli, Parvaneh. 2016. Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching* 54(2). 133–150. DOI: 10.1515/iral-2016-9994.
- Tavakoli, Parvaneh. 2023. Making fluency research accessible to second language teachers: The impact of a training intervention. *Language Teaching Research* 27(2). 368–393. DOI: 10.1177/1362168820951213.
- Tavakoli, Parvaneh & Ann-Marie Hunter. 2018. Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research* 22(3). 330–349. DOI: 10.1177/1362168817708462.
- Tavakoli, Parvaneh, Fumiyo Nakatsuhara & Ann-Marie Hunter. 2020. Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal* 104(1). 169–191. DOI: 10.1111/modl.12620.
- Tavakoli, Parvaneh & Peter Skehan. 2005. Strategic planning, task structure, and performance testing. In Rod Ellis (ed.), *Planning and task performance in a second language*, 239–273. Amsterdam: John Benjamins.
- Terken, Jacques. 1984. The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech* 27(3). 269–289.
- Terken, Jacques. 1991. Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America* 89(4). 1768–1776.
- Terken, Jacques & Julia Hirschberg. 1994. Deaccentuation of words representing 'given' information: Effects of persistence of grammatical role and surface position. *Language and Speech* 37(2). 125–145.

References

- Tolins, Jackson & Jean E. Fox Tree. 2014. Addressee backchannels steer narrative development. *Journal of Pragmatics* 70. 152–164.
- Tottie, Gunnel. 1991. Conversational style in British and American English: The case of backchannels. *English Corpus Linguistics*. 254–271.
- Trofimovich, Pavel & Wendy Baker. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28(1). 1–30.
- Trofimovich, Pavel & Wendy Baker. 2007. Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics* 28(2). 251–276. DOI: 10.1017/S0142716407070130.
- Trouvain, Jürgen & Khiet P. Truong. 2013. Exploring sequences of speech and laughter activity using visualisations of conversations. In *Proceedings of the Workshop on Affective Social Speech Signals (WASSS 2013), Grenoble, 22–23 August, 2013*.
- Tseng, Chiu-yu, Shao-huang Pin, Yehlin Lee, Hsin-min Wang & Yong-cheng Chen. 2005. Fluent speech prosody: Framework and modeling. *Speech Communication. Quantitative Prosody Modelling for Natural Speech Description and Generation* 46(3). 284–309. DOI: 10.1016/j.specom.2005.03.015.
- Tsunemoto, Aki, Rachael Lindberg, Pavel Trofimovich & Kim McDonough. 2022. Visual cues and rater perceptions of second language comprehensibility, accentedness, and fluency. *Studies in Second Language Acquisition* 44(3). 659–684. DOI: 10.1017/S0272263121000425.
- Turco, Giuseppina, Christine Dimroth & Bettina Braun. 2015. Prosodic and lexical marking of contrast in L2 Italian. *Second Language Research* 31(4). 465–491.
- Turk, Alice. 2012. The temporal implementation of prosodic structure. In Marie K. Huffman, Cécile Fougeron & Abigail C. Cohn (eds.), *The oxford handbook of laboratory phonology*, 242–253. Oxford: Oxford University Press Oxford.
- Urbani, Martina. 2012. Pitch range in L1/L2 English: An analysis of F0 using LTD and linguistic measures. In Maria Grazia Busà & Antonio Stella (eds.), *Methodological perspectives on L2 prosody: Papers from ML2P*, 79–83. Padova: Coop. Libreria Editrice Università di Padova.
- Vallduví, Enric. 1991. The role of plasticity in the association of focus and prominence. In *Proceedings of the Eastern States Conference on Linguistics (ESCOL 7)*, 295–306.
- Van Compernelle, Rémi A. & Nuria Ballesteros Soria. 2020. Developing interactional repertoires in the classroom through dynamic strategic interaction scenarios. *Vigo International Journal of Applied Linguistics* 17. 141–169.

- Van Os, Marjolein, Nivja H. De Jong & Hans R. Bosker. 2020. Fluency in dialogue: Turn-taking behavior shapes perceived fluency in native and nonnative speech. *Language Learning* 70(4). 1183–1217. DOI: 10.1111/lang.12416.
- Vander Klok, Jozina Madolyn, Heather Goad & Michael Wagner. 2018. Prosodic focus in English vs. French: A scope account. *Glossa: a journal of general linguistics* 3(1). 1–47.
- Vanrell, Maria del Mar & Olga Fernández-Soriano. 2018. Language variation at the prosody-syntax interface: Focus in European Spanish. In Marco García García & Melanie Uth (eds.), *Focus realization in Romance and beyond*, vol. 201 (Studies in Language Companion Series), 33–70. Amsterdam: John Benjamins Publishing.
- Veilleux, Nanette, Stefanie Shattuck-Hufnagel & Alejna Brugos. 2006. *Transcribing prosodic structure of spoken utterances with ToBI*.
- Verdugo, María Dolores Ramírez. 2003. Non-native interlanguage intonation systems: A study based on a computerized corpus of Spanish learners of English. *ICAME Journal* 26. 115–132.
- Vietti, Alessandro. 2019. Phonological variation and change in Italian. In *Oxford research encyclopedia of linguistics*. Oxford University Press. DOI: 10.1093/acrefore/9780199384655.013.494.
- Visceglia, Tanya, Chiu-yu Tseng, Chao-yu Su & Chi-feng Huang. 2011. Realization of English narrow focus by L1 English and L1 Taiwan Mandarin speakers. In *ICPhS*, 2074–2077.
- Von Heusinger, Klaus. 1999. *Intonation and information structure*. Habilitationsschrift. University of Konstanz.
- Wagner, Petra. 1999. The synthesis of German contrastive focus. In *Proceedings of ICPhS*, vol. 99.
- Waluyo, Budi. 2019. Task-based language teaching and theme-based role-play: Developing EFL learners' communicative competence. *Electronic Journal of Foreign Language Teaching* 16(1). DOI: 10.56040/bdwa161a.
- Ward, Nigel & Wataru Tsukahara. 2000. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics* 32(8). 1177–1207.
- Watson-Gegeo, Karen A. 2004. Mind, language, and epistemology: Toward a language socialization paradigm for SLA. *The Modern Language Journal* 88(3). 331–350.
- Watson-Gegeo, Karen A. & Sarah Nielsen. 2003. Language socialization in SLA. 15(4). 155–171.

References

- Wehrle, Simon. 2023. *Conversation and intonation in autism: A multi-dimensional analysis* (Studies in Laboratory Phonology 14). Berlin: Language Science Press. DOI: 10.5281/zenodo.10069004.
- Wehrle, Simon & Martine Grice. 2019. *Function and prosodic form of backchannels in L1 and L2 German*. Seoul.
- Welby, Pauline. 2004. The structure of French intonational rises: A study of text-to-tune alignment. In *Proceedings of Speech Prosody*, 127–130. DOI: 10.21437/SpeechProsody.2004-30.
- Wenker, Georg. 1877. *Das rheinische Platt: Den Lehrern des Rheinlandes gewidmet*. Im Selbstverlage der Verfassers.
- Wennerstrom, Ann. 1998. Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition* 20(1). 1–25.
- Wennerstrom, Ann. 2006. Intonational meaning starting from talk. In Rebecca Hughes (ed.), *Spoken English, Tesol and applied linguistics: Challenges for theory and practice*, 72–98. London: Palgrave Macmillan UK. DOI: 10.1057/9780230584587_4.
- Wesseling, Wieneke & Rob J.J.H. van Son. 2005. Early preparation of experimentally elicited minimal responses. In *6th SIGdial Workshop on Discourse and Dialogue*.
- White, Sheida. 1989. Backchannels across cultures: A study of Americans and Japanese. *Language in Society* 18(1). 59–76.
- Willett, Jerri. 1995. Becoming first graders in an L2: An ethnographic study of L2 socialization. *Tesol Quarterly* 29(3). 473–503.
- Witton-Davies, Giles. 2014. *The study of fluency and its development in monologue and dialogue*. Lancaster: University of Lancaster. (Doctoral dissertation).
- Wolf, James P. 2008. The effects of backchannels on fluency in L2 oral task production. *System* 36(2). 279–294.
- Wong, Deanna & Pam Peters. 2007. A study of backchannels in regional varieties of English, using corpus mark-up as the means of identification. *International Journal of Corpus Linguistics* 12(4). 479–510.
- Yamamoto, Seiichi, Keiko Taguchi, Koki Ijuin, Ichiro Umata & Masafumi Nishida. 2015. Multimodal corpus of multiparty conversations in L1 and L2 languages and findings obtained from it. *Language Resources and Evaluation* 49(4). 857–882.
- Yngve, Victor H. 1970. On getting a word in edgewise. In *Chicago Linguistics Society, 6th Meeting, 1970*, 567–578.

- Yoon, Tae-Jin. 2010. Speaker consistency in the realization of prosodic prominence in the Boston University Radio Speech Corpus. In *Proceedings of Speech Prosody 2010*. Chicago. DOI: 10.21437/SpeechProsody.2010-278.
- Young, Richard. 2011. Interactional competence in language learning, teaching, and testing. In Eli Hinkel (ed.), *Handbook of research in second language teaching and learning*, vol. 3, 426–443. London: Routledge.
- Young, Richard. 2014. What is interactional competence? In *AL Forum: The newsletter of the TESOL applied linguistics interest section*.
- Zhang, Juan, Yaxuan Meng, Chenggang Wu & Zhen Yuan. 2023. Spoken word recognition across language boundary: ERP evidence of prosodic transfer driven by pitch. *Brain Sciences* 13(2).
- Zimmerer, Frank, Jeanin Jügler, Bistra Andreeva, Bernd Möbius & Jürgen Trouvain. 2014. Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. In *Proceedings of Speech Prosody*, 1037–1041.
- Zmarich, Claudio. 2017. Stuttering and phonetic theory: An introduction. *Studi AISV* 3. 357–372. DOI: 10.17469/O2103AISV000018.

Name index

Albert, Aviad, 17, 37
Allerton, David J., 25
Amador-Moreno, Carolina Pilar, 126
Anderson, Anne, 13, 91, 104
Aronsson, Berit, 2
Arvaniti, Amalia, 38, 75
Ashby, Jane, 1
Austin, Gavin, 8
Avesani, Cinzia, 9, 24, 28, 53, 89, 159
Ayers, Gayle, 74
Aylett, Matthew, 21

Bachman, Lyle F., 94
Baills, Florence, 92
Baker, Wendy, 1, 2, 4, 26
Bard, Ellen Gurman, 21
Baumann, Stefan, 18–21, 25, 33
Beattie, Geoffrey, 100
Beckman, Mary, 20, 74
Beňuš, Štefan, 127
Berger, Evelyne, 102
Bernaisch, Tobias, 128, 129
Berry, Anne, 128, 129
Bertinetto, Pier Marco, 23
Bi, Ran, 27
Bissiri, Maria Paola, 92
Bocci, Giuliano, 24, 28, 53
Bock, Kathryn, 107
Boersma, Paul, 37, 134
Bögels, Sara, 101, 126
Bosker, Hans Rutger, 98
Brazil, David, 18

Broersma, Mirjam, 14, 119
Büring, Daniel, 18, 20
Bürkner, Paul-Christian, 39, 112
Busà, Maria Grazia, 4
Butterworth, Brian, 100

Calhoun, Sasha, 19–21, 33, 87
Campbell, Nick, 20, 102
Campbell-Larsen, John, 2, 8
Canale, Michael, 3, 5
Canepari, Luciano, 23
Cangemi, Francesco, 19, 22, 23, 38, 75, 89, 102, 107
Carpenter, Bob, 39, 112
Castello, Erik, 131
Cekaite, Asta, 2
Chafe, Wallace, 18
Chapple, Eliot, 102
Chodroff, Eleanor, 21
Clancy, Patricia, 127–129, 131
Clark, Herbert, 18
Clingwall, Dion, 8
Cohen, Andrew D., 2, 8
Cole, Jennifer, 21
Crossley, Scott, 163
Cruttenden, Alan, 20–22, 25
Cutrone, Pino, 126, 128, 131

D’Imperio, Mariapaola, 20, 22, 24
Dahmen, Silvia, 92
Darcy, Isabelle, 4
De Jong, Nivja H., 95, 97, 99, 100, 125

Name index

- De Ruiter, Laura E., 21, 90
De Saussure, Ferdinand, 22
Dénes, Mariann, 98
Derwing, Tracey M., 2, 4, 8, 92, 98
Dingemanse, Mark, 102
Dorran, David, 101
Drummond, Kent, 127
Duncan, Hilary D., 96
Duncan, Starkey, 127

Eckman, Fred R., 27, 29, 73, 92
Edlund, Jens, 100, 128, 133
Ellis, Rod, 1

Face, Timothy L., 22
Fang, Wei-Chieh, 160
Farnetani, Edda, 24
Felleggy, Anna M., 127
Fernández-Soriano, Olga, 20
Féry, Caroline, 20, 25, 33
Figueras, Neus, 3, 126
Fillmore, Charles J., 95
Fischer, Kerstin, 92
Fiske, Donald W., 127
Fox Tree, Jean E., 128
Franke, Michael, 40, 112
Freed, Barbara F., 95, 96
Fries, Charles C., 127
Frota, Sonia, 20

Galaczi, Evelina D., 2, 102, 132
Gardner, Rod, 127
Gass, Susan M., 104
Gesuato, Sara, 131
Gili Fivela, Barbara, 23
Gilmartin, Emer, 102
Giordano, Rosa, 23
Gleitman, Lila R., 107
Goad, Heather, 8

Goldman-Eisler, Frieda, 100
Gonzalez, Fernanado, 126, 131
Goodwin, Charles, 127, 128
Gordon, Joshua, 4
Götz, Sandra, 101
Grice, Martine, 18, 21–23, 74, 75, 82, 89, 91, 104, 131
Griffin, Zenzi M., 107
Grosser, Wolfgang, 26
Gussenhoven, Carlos, 19
Gut, Ulrike, 2, 26, 27

Ha, Kieu-Phuong, 126, 129–131
Hadelich, Kerstin, 25
Hahn, Laura, 2, 4
Hall, Joan Kelly, 6, 126
Halliday, Michael, 18–21
Handley, Zoe Louise, 8
Hasegawa, Yoko, 127, 128
Haviland, Susan, 18
Hawkins, Sarah, 7
He, Agnes Weiyun, 6, 94, 95, 126
Heinz, Bettina, 128, 129
Heldner, Mattias, 20
Henderson, Alan, 100
Heritage, John, 127
Hermes, Anne, 25
Hermes, Dik J., 20
Hiligsmann, Philippe, 8, 29, 67
Hirschberg, Julia, 21, 24
Höhle, Tilman N., 27
Hopper, Robert, 127
House, David, 24
Huensch, Amanda, 96
Hunter, Ann-Marie, 8, 96
Hymes, Dell, 2, 5

Ito, Kiwako, 21
Iverson, Paul, 2

- Jacoby, Sally, 5, 126
Jefferson, Gail, 127
Jilka, Matthias, 4
Jung, Yeon Joo, 163
Jurafsky, Dan, 128
- Kahng, Jimin, 101
Kallio, Heini, 1
Kang, Okim, 2, 4
Karttunen, Lauri, 18
Kelly, Niamh, 73
Kendon, Adam, 127
Kim, Eunyi, 29
Kim, Yoo Lae, 14
Kirsner, Kim, 100
Kjellmer, Göran, 128
Klassen, Jeffrey, 19
Kley, Katharina, 162
Kochanski, Greg, 20
Kohler, Klaus J, 25
Kormos, Judit, 1, 6, 96, 98, 99, 107
Kosmala, Loulou, 123
Kousidis, Spyros, 101
Kraaz, Michelle, 128, 129
Krahmer, Emiel, 21, 28, 46, 89
Kramsch, Claire, 2
Kraut, Robert E., 126
Krifka, Manfred, 18, 19
Kügler, Frank, 19, 20, 25, 33, 87
Kuno, Susumu, 19
Kurumada, Chigusa, 21
- Ladd, D. Robert, 19, 20, 22, 23, 26, 28, 75
Lahmann, Cornelia, 96
Lambertz, Kathrin, 127
Lambrecht, Knud, 18
Leiner, Dominik J., 47
Lemhöfer, Kristin, 14, 119
- Lemoine, Nathan P., 79
Lengeris, Angelos, 4
Lennon, Paul, 95, 96, 127
Lepschy, Anna Laura, 23
Lepschy, Guilio, 23
LeVelle, Kimberly, 4
Levinson, Stephen C., 7, 101, 107, 126
Levis, John, 4
Lewis, David, 18
Li, Han Z., 126, 128, 130
Liesenfeld, Andreas, 102
Lintunen, Pekka, 101
Loewen, Shawn, 8
Logan, Kenneth J., 96
Long, Mike, 104
Loporcaro, Michele, 23
- Macdonald, Shem, 8
Mackey, Alison J., 104, 160
Magne, Viktoria, 98
Magno-Caldognetto, Emanuela, 23
Major, Roy C., 29
May, Lyn, 102
McCarthy, Michael, 101, 126
McDonough, Kim, 14, 162
Mennen, Ineke, 2, 72
Missaglia, Federica, 92
Morrison, Astrid, 8
Mücke, Doris, 21, 22
Munro, Murray J., 2, 4, 8, 92
- Nakane, Ikuko, 2
Nava, Emily A., 8, 122
Nicora, Francesca, 92
Niebuhr, Oliver, 23, 92
Nielsen, Sarah, 1
North, Brian, 2, 5
Nurjaleka, Lisda, 128
- Ochs, Elinor, 5, 126

Name index

- Palmer, Adrian S., 94
Paul, Hermann, 12
Pekarek Doehler, Simona, 102
Peltonen, Pauliina, 6, 93, 95, 100–102
Pérez, Nubia Patricia Carrero, 160
Peters, Pam, 128
Petrov, Sergei V., 73
Piccardo, Enrica, 2, 3, 5–7, 94, 101, 104, 126
Pierrehumbert, Janet, 20, 74
Pika, Simone, 7
Pillai, Stefanie, 26, 27
Piontek, Jörn, 21
Piske, Thorsten, 2
Préfontaine, Yvonne, 98
Prieto, Pilar, 20
Prince, Ellen F., 18
Ptak, Alexander, 102
Purcell, Edward T., 92

Rasier, Laurent, 8, 26, 29, 67
Ravignani, Andrea, 7
Refice, Mario, 128
Reinhardt, Susanne, 8
Riester, Arndt, 18, 21, 25
Riggenbach, Heidi, 97, 100, 101
Ritter, Simon, 75
Rivers, Wilga M., 1
Roberts, Benjamin, 100
Roberts, Felicia, 7
Roessig, Simon, 22
Roettger, Timo B., 21, 40, 112
Röhr, Christine Tanja, 21
Rossiter, Marian J, 98
Rühlemann, Christoph, 102, 127
Rump, H. H., 20

Sacks, Harvey, 101, 126
Saito, Kazuya, 97–99

Samek-Lodovici, Vieri, 20
Sato, Masatoshi, 8, 101
Savignon, Sandra J., 3
Savino, Michelina, 23, 104, 128, 133, 135, 148, 162
Sbranna, Simona, 2, 8, 17, 31, 37, 93, 103, 104, 125, 126, 162
Schegloff, Emanuel A., 7, 126, 127
Schiffer, Stephen R., 18
Schnur, Tatiana T., 107
Schweitzer, Katrin, 21
Segalowitz, Norman, 1, 95–97, 100, 104
Selinker, Larry, 25
Selkirk, Elisabeth, 20
Senk, K. Maynard, 128
Shelley, Leah, 126, 131
Shi, Shuju, 72
Shibata, Miki, 162
Simon, Carsta, 127
Skehan, Peter, 1, 96, 97
Sluijter, Agaath M. C., 20
Smorenburg, Laura, 92
Song, Jieun, 2
Sørensen, A. Josefine Munch, 2, 102
Soria, Nuria Ballesteros, 2
Spaniol, Malin, 102, 162
Stalnaker, Robert, 18
Starkweather, C. Woodruff, 96
Stechow, Arnim, 18
Stivers, Tanya, 7
Stocksmeier, Thorsten, 130
Suter, Richard W., 92
Suzuki, Shungo, 98
Swain, Merrill, 3, 5
Swerts, Marc, 9, 21, 27–29, 46, 89, 159

Takahashi, Daniel Y., 7
Takeda, Aya, 27

- Tao, Hongyin, 128
Tavakoli, Parvaneh, 1, 6, 8, 96–98, 101
Terken, Jacques, 19–21
Thompson, Sandra A., 128
Tolins, Jackson, 128
Torreira, Francisco, 7, 101, 107, 126
Tottie, Gunnel, 128, 129, 135
Tracy-Ventura, Nicole, 96
Trofimovich, Pavel, 1, 2, 4, 26
Trouvain, Jürgen, 102
Truong, Khiet P., 102
Tseng, Chiu-yu, 1
Tsukahara, Wataru, 129
Tsunemoto, Aki, 162
Turco, Giuseppina, 27
Turk, Alice, 20

Urbani, Martina, 72

Vallduví, Enric, 21
Van Compernelle, Rémi A., 2
Van Heuven, Vincent J., 20
Van Os, Marjolein, 2
van Son, Rob J.J.H., 107
Vander Klok, Jozina Madolyn, 19
Vanrell, Maria del Mar, 20
Vayra, Mario, 24
Veilleux, Nanette, 74
Verdugo, María Dolores Ramírez, 26, 27
Vietti, Alessandro, 23
Visceglia, Tanya, 72
Von Heusinger, Klaus, 18

Wagner, Petra, 25
Waluyo, Budi, 160
Ward, Nigel, 129
Watson-Gegeo, Karen A., 1
Weenink, David, 37, 134

Wehrle, Simon, 7, 131, 136, 148
Welby, Pauline, 76
Wenker, Georg, 12
Wennerstrom, Ann, 1, 26
Wesseling, Wieneke, 107
White, Lydia, 8
White, Sheida, 130
Willett, Jerri, 1
Winter, Bodo, 20
Witton-Davies, Giles, 6
Wolf, James P., 127
Wong, Deanna, 128

Yamamoto, Seiichi, 2
Yngve, Victor H., 126, 127
Yoon, Tae-Jin, 21
Young, Richard, 1, 5, 6, 94, 95, 126

Zerbian, Sabine, 27
Zhang, Juan, 8
Zimmerer, Frank, 72
Zmarich, Claudio, 24, 96
Zubizarreta, Maria Luisa, 8

Prosody and interactional fluency of Italian learners of German

This book explores the development of prosodic and interactional competence in second language acquisition, drawing on data from peer interactions by Italian learners of German in both German and their native language, Italian, as well as from German native speakers. Three key aspects of spoken interaction are examined across proficiency levels: prosodic marking of information status, turn-taking, and backchannels. The analysis of prosodic marking of information status reveals that learners mark givenness using distinct fundamental frequency patterns, as in their native language, but apply a reduction in prosodic strength typically found postfocally in native German, irrespective of its function. This suggests that learners perceive deaccentuation as a salient marker of native German, which they adopt during their learning. This book also presents a novel approach to quantifying interactional competence, showing that lower proficiency negatively affects the smoothness of interactional flow, resulting in reduced speech time and increased overall silence. Finally, it provides new insights into backchannel use in second language and cross-linguistic contexts. Results show a complex, non-arbitrary mapping between lexical type, turn-taking function, and intonation in both native languages. In second language speech, dyad-specific behaviour was found to have a stronger effect on backchannel frequency and duration than second language proficiency. Furthermore, learners tend to transfer preferred lexical backchannel types from their first language into their second language. Overall, this book offers a multidimensional perspective on second language spoken interaction and lays the groundwork for future applications in language teaching and assessment.

The doctoral work, on which this book is based, was awarded the IPA PhD Thesis Award for the “Best PhD Thesis in the broad area of Phonetics, Speech Sciences, and Laboratory Phonology” in 2024.