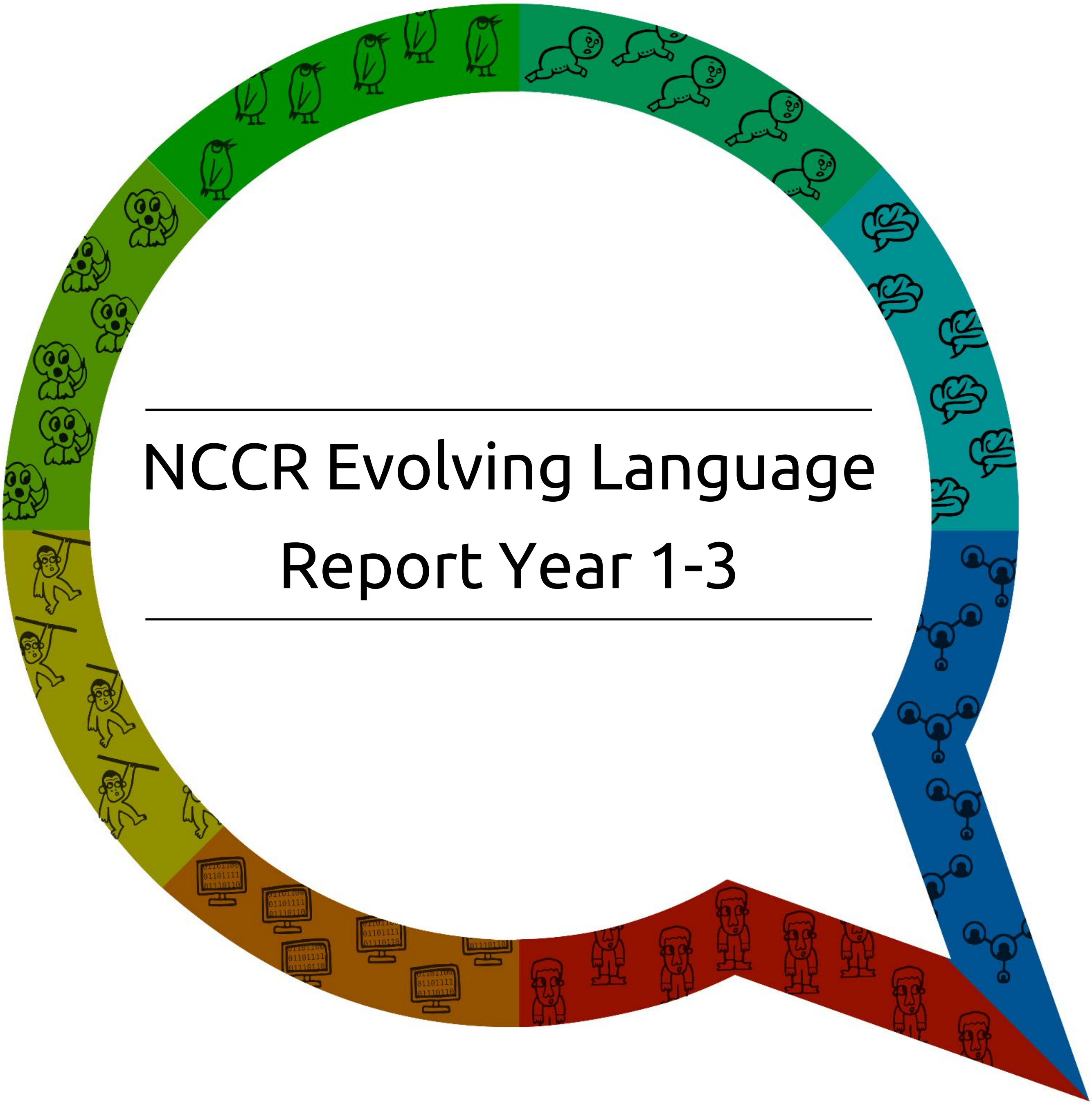
# NCCR Evolving Language Report Year 1-3



Daphné Bavelier, Klaus Zuberbühler, Simon W. Townsend, Emilie Wyss & Balthasar Bickel

*July 2023*

The National Centres of Competence in Research (NCCRs) are a funding scheme of the Swiss National Science Foundation.

Cite as:
Bavelier D, Zuberbühler K, Townsend SW, Wyss E & Bickel B (eds). 2023. Report Year 1-3 of the NCCR Evolving Language. Zurich/Geneva: NCCR Evolving Language.
https://evolvinglanguage.ch/wp-content/uploads/NCCR-Report-Y3-.pdf

# Executive Summary

The grand vision of the NCCR is to unravel the evolutionary origins of language so we can cope with various 21st century challenges from how cultural evolution impacts language, with rapidly evolving clinical technologies, a fast growing presence of artificially intelligent communicators and steep declines of linguistic diversity and identity.

To this end, we have expanded the Tinbergian toolkit of evolutionary research beyond biology. In terms of proximate mechanisms and their ontogenies, we have placed emphasis on the fact that the language phenotype is a fundamentally dynamic trait, following its own distinct mode of linguistic evolution, and on the fact that technological evolution allows us to increasingly intervene in core language functions and their ontogenetic trajectories, with promising clinical potential and substantial ethical challenges. In terms of ultimate phylogeny and adaptive value, we have focused on the social context as a key driver in the emergence of increasingly more complex communication systems, and we have highlighted language models as a novel kind of interlocutor, driven by, and driving further, cultural evolution in our societies.

In terms of overall findings, several studies point to the same emerging conclusion, i.e., that many components of language, such as non-adjacent dependency processing, the propensity to conceptualise events in terms of semantic roles, or oscillatory behaviour in vocalisations, are in fact not uniquely human but rather build on evolutionarily older computational capabilities shared with several other species. The same conclusions appear to emerge from research on social behaviour — what has so aptly been called the human interaction engine, a prerequisite for language, appears to be present in other species where it serves the same function, i.e., to ensure social bonding and to provide a launchpad for prosociality and joint action. By contrast, our research in and on technology suggests that language models are still far away from neurobiologically realistic machines, both in terms of computational formats and fits with predicted neural activity, even if their output is difficult to distinguish from human output. At the same time, we showed that technologies are increasingly successful in intervening on core language functions, for example, through brain-computer interfaces, neuromodulation or attention boosting.

In the structure-related areas, our main achievements over the past three years are in fostering inclusivity, both through programmes ensuring equal opportunities and recognition of diversity and in terms of education activities with partner institutions from socio-economically disadvantaged regions. Another achievement we wish to highlight is proactive science communication, through social media, outreach events, educational resources and close collaboration with traditional media channels. A further achievement of note is that we have developed an efficient pipeline for sustainable data management, which helped us push open research data at all levels.

Finally, the structural measures of our host institutions are by now all fully implemented even though filling the last two of the planned professorships (genetics and paleoanthropology) might be delayed by a few months. Together, the implemented measures put us in an excellent position for Phase 2, with much new expertise and a rich set of results to capitalise on.

## Results from Years 1-3: Highlights and output

From a classical Tinbergian perspective, four questions need to be answered before we can model the origins of the language faculty. A first 'proximate' set is concerned about (1) the computational and neurobiological mechanisms that define this faculty and (2) how they develop ontogenetically. A second 'ultimate' set is how these traits (3) function in terms of their adaptive value and (4) how they have evolved phylogenetically. Although originally developed for biology, the Tinbergian framework is equally applicable to cultural phenotypes evolved over historical time, including language.

To answer the 'proximate' questions, we have focused on the fact that language is a fundamentally dynamic trait, with its own mode of linguistic evolution and acquisition mechanisms, and with one of the most complex temporal processing dynamics the human brain has to face. With this focus, we moved beyond traditional approaches that have emphasised static and algebraic properties of the language faculty when characterising it as a phenotype that has evolved in the hominin lineage. At the same time we moved beyond biology and examined the impact of cultural, in particular technological, evolution as a force that intervenes on mechanisms (e.g. through circumventing the auditory-phonetic loop through decoding of syntactic plans in the brain) and their ontogenetic development (e.g. through early interventions on developmental trajectories in language skills).

To answer the 'ultimate' questions, we have focused on targeted contrasts between species that identify which aspects of language are shared with other animals by common descent or convergence. We moved beyond traditional approaches that emphasise human uniqueness by trying to unravel a complex web of traits that are shared to varied degrees and that reflect possibly distinct evolutionary trajectories. Also, as in the case of proximate questions, we moved outside biology and put emphasis on how cultural evolution (e.g. in social structures or technology) might have impacted the emergence of language and its neural bases, and how such impact might continue in the future, in particular with the advent of language models as new interlocutors in our societies (e.g. chatGPT).

In the past three years, each of our nine projects has worked on both sets of questions simultaneously, although emphases naturally differed. By doing so, the teams have grown together, with "proximists" developing increasing interest in comparative questions, and "ultimists" developing an increasing interest in mechanistic questions. For example, in the Cooperation Project, psycholinguists and primatologists started to compare the relevance of immature-directed speech across primates, in the Grammar Project, linguists and primatologists probed the agent preference across hominids, or in the Reception Project, neuroscientists and phylogenetic modellers examined the distribution of theta-band frequency in vocalisations across vertebrates, putting to the test the human-centred hypothesis of theta-band frequency as a key level of neural organisation for vocal communication, just to highlight a few examples.

In what follows, we report selected findings in each project, some illustrated graphically and listing all publications in tabular form with information on open science as requested by the SNSF. Note that many projects were fully staffed only by 2021 and that in many cases research was heavily delayed by the pandemic; we therefore also list publications that are still under review or revision. For more detailed and more comprehensive reports on each project, please see our Y2 Progress Report.

In order to highlight the transdisciplinary nature of the WPs, we added a colour scheme tracking the main disciplines in the NCCR (mirroring what we have on our website, www.evolvinglanguage.ch):

Table 1. Colour code representing the different disciplines in the NCCR.

| | | | |
|---|---|---|---|
| Linguistics | Computer Science and Mathematics | | Biology and Anthropology |
| Neuroscience | Psychology | Medicine | Philosophy and Ethics |

## 1 Grammar Project

The overarching goal of the Grammar Project in Phase I is to chart the similarities and the differences between humans and non-human animals across three fundamental aspects of grammatical mechanisms: (a) *compositionality*, i.e., the phenomenon where the meaning of a combination derives not only from the meaning of the parts, but also from the specific way that the parts are combined, (b) *semantic role structure*, i.e., the fact that human languages decompose events into a core event and the participants that play a role in this event and, lastly, (c) *emotion integration*, i.e., the notion that propositional information is integrated with emotional information at the utterance level.

In the **WP Compositionality** we have applied data-driven stochastic modelling approaches to reveal a hitherto underappreciated level of internal structuring in the call sequences of non-human primates [1]. These findings challenge the long-held view that animal vocal sequences only bear superficial similarities with human syntactic structures. At the computational level, we find that causally compositional data sets created via generative models remain a challenge for existing artificial communicators based on neural network models. This suggests a gap in how humans and machine process language and suggest that more customised architecture is needed if compositionality is to be fully captured [2]. *PIs: Bickel, Merlo, Townsend; Collaborating PIs: Burkart, Stoll, Zuberbühler; Staff: An, Bosshard*

The **WP Semantic Roles** has shown that a default agent interpretation in language comprehension and gist apprehension is resilient against cross-linguistic variation and independent of communicative experience [3]–[6]. Comparative work (adults and children, chimpanzees, gorillas and orangutans) highlights a similar preference for agents when acting on inanimate objects [7] (Figure 4), specifically food [8], a preference that is not reducible to general causation attribution [9]. Such convergent results point towards shared ecological contexts being a key driver in how event cognition has shaped human syntax. *PIs: Bickel, Daum, Zuberbühler; Collaborating PIs: Merlo, Stoll, Townsend; Staff: Wilson, Huber, Pascual, Brocard; Internal Collaborators: Andrews, Isasi-Isasmendi, Schnell, Wermelinger; External Collaborators: Itziar Laka (U Vitoria-Gasteiz), Ina Bornkessel-Schlesewsky (U. South Australia), Monique Flecken (U Amsterdam), Marina Kalashnikova, Simona Mancini (both BCBL, San Sebastian); Associate Investigator: Sauppe*

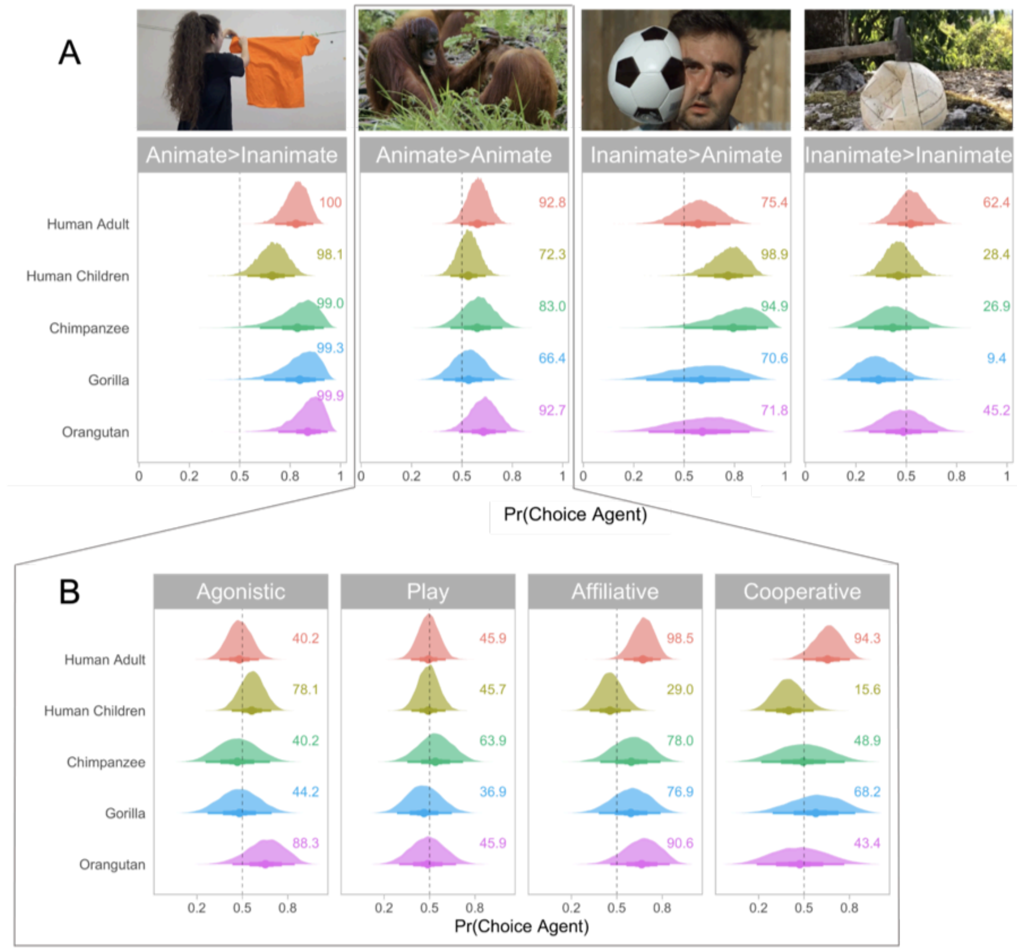



Figure 4. All great apes preferentially select agents when acting on inanimate patients if given the choice on a touch screen after seeing a video clip of dyadic interactions. Beyond this general pattern, we found additional species-specific biases, such as when adult humans react to cooperative social interactions or when chimpanzees and human children react to inanimate agents acting on animate patients, a typical scenario during tool-use. Overall, our findings point to highly articulated, specific mechanisms in hominid cognition that go beyond the general attentional bias towards self-propelled entities and natural causes that are attested in many animals [10]. The findings echo a prior in language processing where initial, role-ambiguous noun phrases are transiently assigned an agent role, a prior that is most cross-linguistically strongest when agents are animate [11], [12].

Lastly, in the **WP Emotion Integration** human behavioural studies investigating the relationship between semantic and prosodic emotional aspects in the perception of sarcasm and irony indicate that sarcasm is most salient when a negative context and statement is coupled with a positive prosody, whereas irony is most salient when a negative context is coupled with a positive statement and prosody. Interestingly, parallel research investigating the interplay between emotional and contextual information on meerkat contact vocalisations supports a dominant role of behavioural context over arousal (a proxy of emotion) in explaining acoustic variation in this core call type. Through highlighting the continuities but also the discontinuities between animals and humans these results pave the way to disentangle the trajectories of who different types of information are communicated. *PIs: Grandjean, Manser, Xanthos; Collaborating PI: Merlo; Staff: Driscoll, Mayr, Wittmann, Gupta, Benkais; External Collaborators: Sascha Fruehholz (U Oslo), Adrien Meguerditchian (U Aix-Marseille); Elodie Briefer (Copenhagen)*

## 2 Diversification Project

The Diversification Project seeks to capture the spatio-temporal dynamics of communication in humans and animals, a prerequisite for charting the transition from biologically fixed systems to the linguistic evolution that characterises human communication.

In the **WP Arbitrariness** we revisited the long-standing dichotomy that human language is arbitrary in its form-meaning mapping whilst animal communication is hardwired, with vocalisations in particular being irrevocably tied to the arousal state or behavioural context experienced by the signaller. Through devising a novel multi-layered conceptual framework that focuses more on optionality and on the capacity to associate alternative functions with a signal (Figure 5), we provide a powerful new approach allowing cross-species investigation of the origins and evolution of flexible form–function mappings [13]. *PIs: Manser, Widmer; Collaborating PIs: Townsend, van Schaik, Zuberbühler; Staff: Tamer, Falk, Watson*

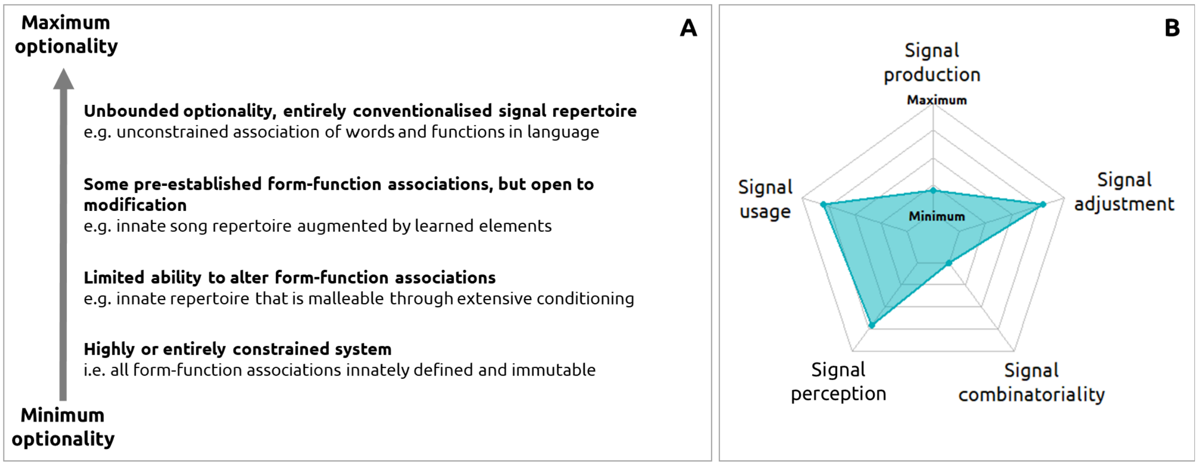



Figure 5. (A) Spectrum of optionality with examples of each degree. The four degrees highlighted are not intended to be a prescriptive, quantitative evaluation of points on this spectrum but rather an aid for conceptualising our proposed framework. (B) Plot representing five axes of optionality in the communication system of a hypothetical species, illustrating that a species may demonstrate a high degree of optionality in one domain but not another.

**WP Phylo** has developed a suite of methods to better disentangle horizontal from vertical signals [14], [15] and speed of change from long-term preferences in linguistic evolution [16] as well as to explore the extent to which patterns of linguistic evolution covary with patterns of phylogeographic dispersal [17]–[19]. Through combining genetic and linguistic information on a global scale GeLaTo (Genes and Languages Together, 7) we further showed that populations and languages evolve largely in tandem but that shifts are regular and common outcomes as well [21], [22]. *PIs: Shimizu, Weibel, Widmer; Collaborating PIs: Bickel, Furrer, Sánchez-Villagra (until 2022); Staff: Neureiter, Barbieri, Cathcart, Ranacher*

## 3 Computation Project

The Computation Project aimed to understand the language faculty from a computational perspective that takes to heart the temporally structured nature of its biological substrates. This is a key precondition for understanding how the language phenotype might have evolved from more widespread neuro-computational mechanisms.

In **WP Neurocomputation** we have probed the relative contribution of neural oscillations and predictions in the processing of the continuous speech stream. Findings from the last three years suggest that a combination of theta-gamma coupling and predictions [23] are central to segmenting the speech sound wave into its constituent discrete syllables. In addition, beta oscillations in particular provide a useful mechanism to combine internal predictions with sensory information in a flexible way [24]. Most recently, we have been exploring the computational principles used by humans when faced with semantically ambiguous words. Our findings suggest, in contrast to Large Language Models, which rely on next word predictions, human neuronal signatures are better explained by structured predictions organised in a temporal hierarchy (Figure 6) [25]. *PIs: Giraud, Van De Ville; Collaborating PIs: Bourlard, Hahnloser, Henderson; Staff: Su, Olasagasti; External Collaborators: Lucy MacGregor (U of Cambridge)*

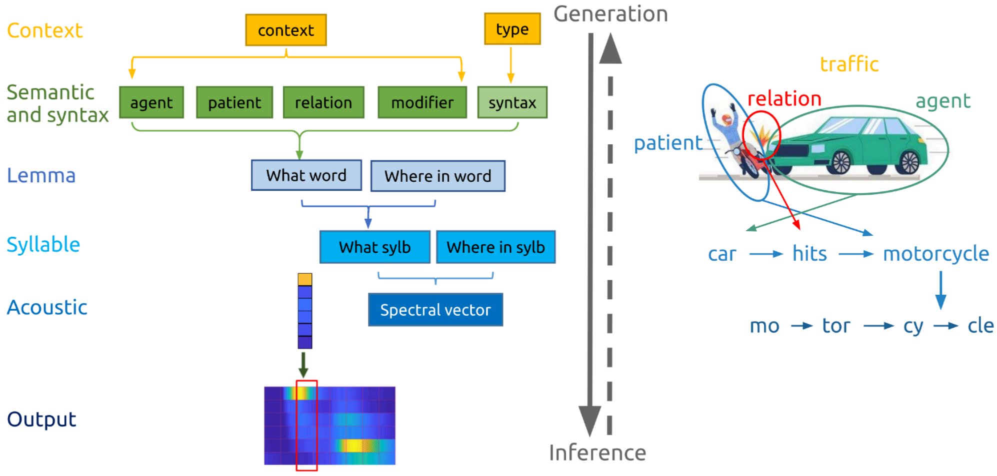



Figure 6. Schematic of a generative model of speech. Left: information conveyed in a spoken sentence is roughly separated into six hierarchies from the situational context to the continuous acoustic signal. Elements in the superordinate level determine those in the subordinate levels in a probabilistic fashion ("predictions"). Inference corresponds to the probabilistic inversion of the generative process. Right: cartoon illustrating how a sequence of syllables ‘/ˈməʊtəˌsaɪkəl/’ (motorcycle) is generated from a traffic scene context. Importantly, the same lemma can represent different semantic roles under different contexts, and so context information is needed for semantic disambiguation.

**WP SignalRec** have implemented computational modelling to better understand the processing role of the cochlea in speech comprehension in addition to unpacking the mechanisms required to capture the structure of language. Results indicate (i) that relatively simple computational models are sufficient to identify not only macroscopic structures present on the cochlea but also those known to exist in higher levels of cognitive processing and (ii) that attention-based architectures, such as Transformers approximate nonparametric Bayesian inference of Dirichlet processes [26]. We furthermore devised a novel nearest-neighbour algorithm, ground-truthed on existing speech corpora, to allow the retrieval of acoustically similar vocalisations from a continuous vocal stream. This approach holds great potential to revolutionise the field of comparative communication, helping to automate data processing and extraction, but also opening up a range of research questions regarding acoustic similarity, that were previously out of reach. *PIs: Bourlard, Hahnloser, Henderson; Collaborating PIs: Garner, Giraud, Merlo, Van De Ville, Stoll; Staff: Fehr, Coppieters*

## 4 Reception Project

In the Reception Project we investigated the neural underpinnings of speech processing through focusing on the role of temporal windows of analysis and the contribution of prosodic information in acoustic-to-linguistic parsing. By doing so, the project identifies candidates of phylogenetically old rhythmic mechanisms and separates these from what has evolved specifically with and for language.

**WP TemporalHierarchies** have revisited the uniqueness of the theta range for speech perception through investigating whether audio-vocal convergence around the theta rhythm also occurs in a diverse range of species. Collating vocal sequence data from a total of 98 species, we show a clear cross-species bias towards a narrow rhythm frequency of approximately 3KHz suggestive of shared temporal reception constraints (Figure 7) [27]. A parallel, targeted study focusing on speech-rhythm tracking in dogs mirrors the cross-species findings, suggesting a perceptual tuning towards a 3Hz window of analysis [28]. *PIs: Giraud, Grandjean, Meyer; Collaborating PIs: Bickel, Dellwo; Staff: Déaux, Van Ommen, Piette; External Collaborators: Adrien Meguerditchian (U Aix-Marseille), Luc Arnal (Institut Pasteur)*

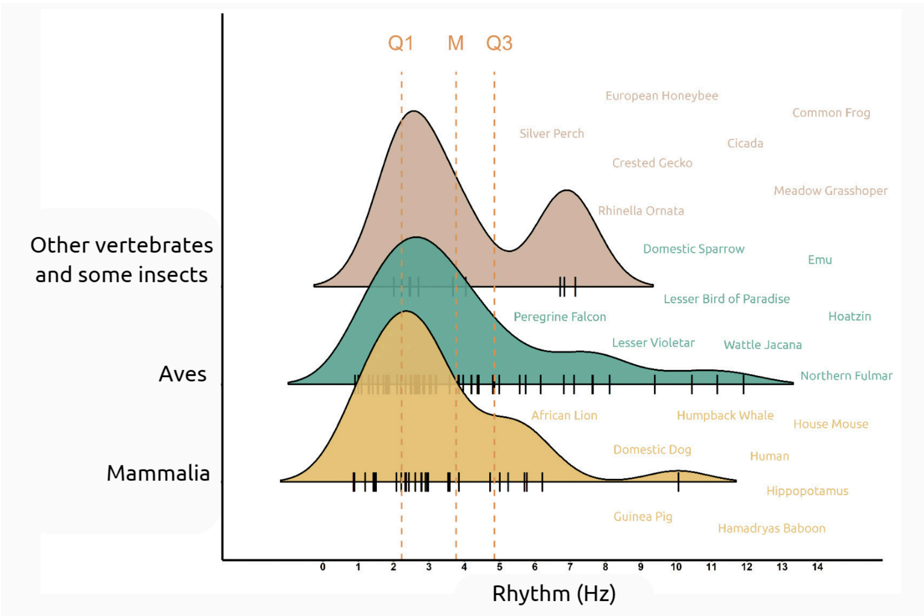



Figure 7. Density distribution of vocal rate (Hz) across clades. Black bars within clades represent species averages and dashed lines show overall median and interquartile ranges. Named species are examples from each group to give a sense of the diversity.

In **WP ProsodyToMeaning** neural measurements of oscillatory brain responses during speech processing indicate (i) a dominant neural tracking at the syllable and intonation level [29] and (ii) prosodic modulation of the speech signal boosts the neural processing of syntax [30]. Together, this research suggests that the faster segment-level chunking characteristic of human speech processing may well be derived in the hominid lineage and thus unique to humans, but also emphasises the importance of slower rhythms in spoken language in both low-level and high-level linguistic processing. *PIs: Golestani, Meyer; Collaborating PIs: Bickel, Giraud, Grandjean, Merlo, Townsend; Staff: Degano, Oderbolz; External Collaborators: Laura Gwilliams (Stanford), Peter Donhauser (Strüngman Institute), David Poeppel (NYU and Strüngman Institute); Associate Investigator: Sauppe*

## 5 Production Project

The Production Project aimed to gain deeper insight into various key stages of language production, from initial early-stage planning to subsequent integration of motor planning and articulation of speech motor gestures. Our work is beginning to shed light on the neural bases of the production of hierarchical syntactic structures, a prerequisite for exploring their evolution, in addition to clarifying when, from a motor-speech planning perspective, the adult-like state emerges and the impact of anatomical properties on speech articulation.

**WP StructurePlanning,** in particular, focused on the early stages of language production, namely how hierarchical structures are deployed in the generation of utterance plans. Through exploiting languages with different case-marking dependencies (Swiss German, Hindi, Basque) we showed for the first time that event-related changes in theta, alpha and beta power index early syntactic planning processes [31], [32] and we are now probing this deeper in cross-serial vs nested and adjoined dependencies. Applying temporal response function analysis to spontaneous speech production, we have furthermore identified the syntactic features that best predict cortical activity in intracranial sEEG recordings [33] (Figure 8). *PIs: Bickel, Meyer, Giraud; Collaborating PIs: Golestani, Henderson, Laganaro, Merlo; Staff: Shaw, Morucci; Associate Investigator: Sauppe*

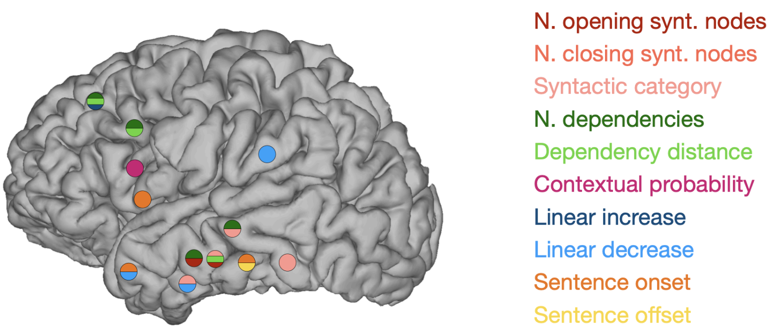



Figure 8. Predicting cortical activity during spontaneous speech production from diverse syntactic properties. Each circle on the surface represents an electrode contact, coloured by the properties that significantly improved prediction accuracy compared to a baseline model including only word-level properties without syntactic context (frequency, word onset, semantic vector)

The focus for the **WP SpeechArticulation** was on later stages of language production, investigating the ontogeny of speech production and the role of articulatory anatomy in explaining potential differences in speech production. Findings from two parallel developmental studies suggests that, firstly, from an articulation perspective, adult-like speech competence emerges from adolescence onwards [34]. Secondly, through measuring ERPs, adolescents and adults were found to have increased ERP difference when planning speech versus matched non-vocal oral gestures, whereas such a difference was absent in children [35]. Regarding anatomical influences on articulation, we found an effect of jaw length on syllable frequency, but only when producing fast speech, an effect that seems likely driven by shorter mouth closure movements in people with smaller mandibles [36]. The fact that such an effect was not found during relaxed speech suggests behavioural accommodation to common rhythmic patterns. *PIs: Dellwo, Laganaro; Collaborating PIs: Giraud, Grandjean; Staff: Lancheros, Atanasova, Friedrichs; External Collaborators: Stuart Rosen (UCL), Michael T. Johnson (U. Kentucky), Francis Nolan (U. Cambridge)*

## 6 Variability Project

Since Darwin, it has been clear that the key data for reconstructing evolution is variation. Language behaviour exhibits considerable variability both across (e.g. multilingualism) and within (across the lifespan) individuals, yet the extent of the variability and its origins remain at large. The Variability Project aimed to bridge this gap in understanding.

In **WP Aptitude** we have addressed variation by coupling linguistic and cognitive behavioural skills assessment with measures of brain function and structure (fMRI, structural MRI, myelin mapping and DTI) in individuals varying in (multilingual) language background and competence. Results suggest that language performance measures are associated with general cognitive skills, in line with other recent work [37], but that inter-individual differences in aptitude – due to their lack of association with multilingual language experience, may arise at least in part from additional genetic predisposition factors. Neuroimaging conducted within this WP also highlighted brain structural differences as a function of multilingualism and language aptitude. Namely, greater multilingual language experience with greater phonological diversity was associated with greater differences in the cortical thickness of the transverse temporal gyri [38]. *PIs: Berthele, Golestani; Collaborating PIs: Brem, Daum, Meyer; Staff: Rampini, Balboni; Internal Collaborators: Kepinska, Maghsadhagh, Udry; External Collaborators: Anja-Xiaoxing Cui (U Vienna,), Christian Doeller (MPI-CBS Leipzig), Evelina Fedorenko (MIT), Simon Fisher (MPI Nijmegen), Clyde Francks (MPI Nijmegen), Jutta Mueller (U Vienna), Tomas Paus (St Justine), Robert Zatorre (McGill), Cathy Price (UCL)*

In the **WP Neuromodulation** we aimed to apply neuro-interventional measures (e.g. neuromodulation, neurofeedback) to elucidate how to act on the neural bases of varied language-related skills. Using transcranial electrical stimulation (tES) techniques and targeting network interactions rather than just selective brain regions, we could demonstrate a robust improvement in word learning in stroke patients [39]. The performance of our brain-computer interface (BCI) for syllable decoding in adults improved over five consecutive days (Figure 9B). These findings will help us to characterise neural mechanisms involved in operating speech-BCI and provide deeper insights into language functions. Furthermore, using real time fMRI neurofeedback, we have shown it is possible to train participants to upregulate activity in their Visual Word Form Area (VWFA) and that this was accompanied with increased activation across the whole reading network (Figure 9C) [40], thereby suggesting deeper causal mechanisms that underlie variation in language functions. *PIs: Brem, Guggisberg; Collaborating PI: Giraud; Staff: Frei, Haugg, Marchesotti, Farcy, Bhadra*

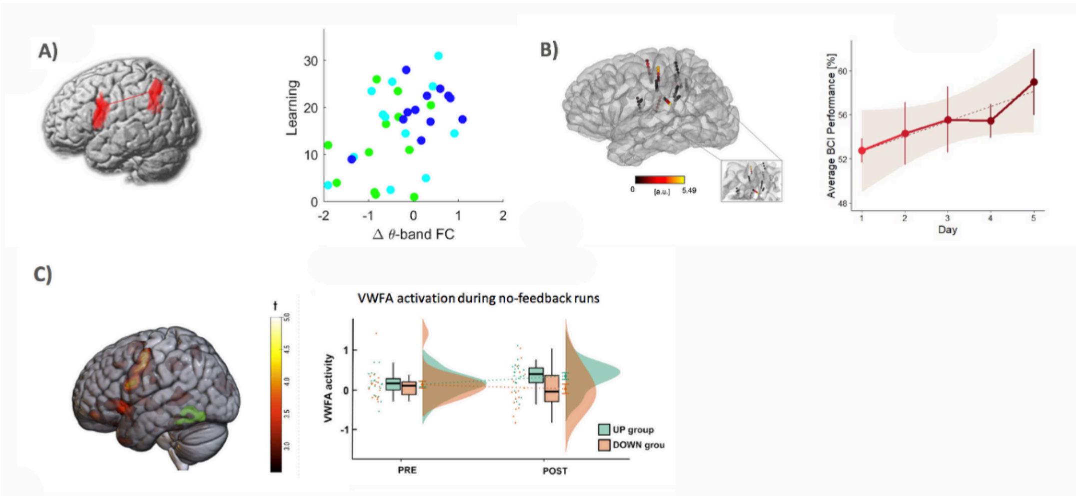



Figure 9. A) transcranial direct current stimulation (tDCS) over the inferior frontal gyrus enhanced interactions (functional connectivity) with the temporo-parietal junction (left); this network enhancement correlated with better learning (right); B) Intracranial-BCI electrodes that best discriminate between the two imagined syllables in one patient (n=1, left); BCI-control performance in healthy volunteers improves over the course of 5 days of training (n=15, right). C) Activation when comparing the upregulation (n=20) to the downregulation (n=20) group during neurofeedback regulation runs (left); activation levels in the Visual Word Form Area (green area) before (pre) and after (post) neurofeedback training (right).

With **WP DigitalDisorder** we show how state-of-art immersive virtual reality can augment word-learning abilities in healthy participants and soon in stroke patients [41]. Furthermore, we have documented how fMRI-based real time neuro-feedback methods when applied to participants with developmental dyslexia can be used to reveal alterations in brain connectivity during letter-speech sound learning [42]. This work highlights once more that the evolution of the relevant mechanisms are to be characterised in terms of networks and connections, and not individual brain regions, a challenge to be taken up in the future. At the same, our findings suggest new routes towards clinical applications. *PIs: Brem, Laganaro; Collaborating PIs: Bavelier, Guggisberg, Zesiger; Staff: Frei, Haugg; Internal Collaborators: Röthlisberger; External Collaborators: Karin Landerl (U Graz), B. Glize (CHU Bordeaux), Ulla Richardson (U Jyväskylä), Milene Bonte (U Maastricht)*

## 7 Sociality Project

Communication is an inherently social endeavour and language is no exception. While the role of the social world in shaping communication phylogenetically and ontogenetically has received considerable attention, the influence of social interaction on communicative competence remains less clear. We bridged this gap in understanding through investigating the relative contribution of surrounding and directed communication and the influence of the accompanying communicative context on communicative competence in humans and our closest-living relatives, the great-apes.

In **WP EarlySurround** we have highlighted the potentially derived nature of child-directed communication in humans [43] and validated this with observational fieldwork in our closest living ape relatives, bonobos and chimpanzees: immature-directed vocal communication in apes is exceedingly rare. We propose that during hominin evolution the importance of child-surrounding communication gradually decreased with the opposite trend for child-directed communication (Figure 10). Furthermore, with a cross-cultural preference experiment, we have implicated a role for child speech (in addition to child-directed and child-surrounding communication) as a previously neglected yet salient input. *PIs: Stoll, Zuberbühler, Van Schaik; Collaborating PIs: Burkart, Townsend; Staff: Fryns, Wegdell, Schick; Internal Collaborator: Lara Nellissen*

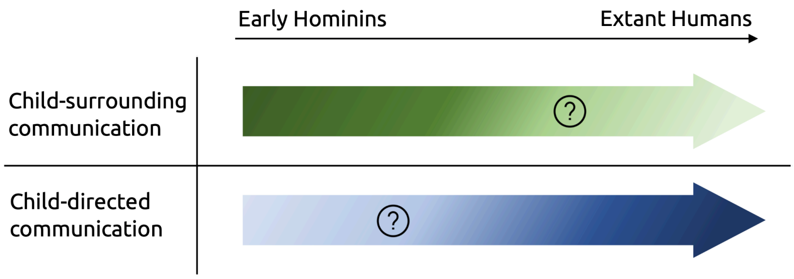



Figure 10. Illustration of two opposing trends in hominin evolution. In the ancestral state, infant hominins were/are exposed to little directed communication to the effect that they had to acquire their communicative competence exclusively from attending to the surrounding environment. Modern human infants, in contrast, benefit much from directed communication, which much accelerates learning.

In **WP SocialContext** our goal has been to chart the development and presence of extra-linguistic cues (ELCs body postures, facial expressions, gestures) and their role in meaning disambiguation in children and chimpanzees respectively, but also from a computational perspective. To this end we have (i) adapted visual gaze-detection algorithms to systematically assess gaze behaviour and what cues are being attended to during utterance production [44], (ii) built the first repertoire of ELCs in any non-human species, identifying a potential role for learning in the acquisition of this repertoire [45], and (iii) developed the first methods for the unsupervised induction of meaningful units in character sequences [46]. One recent finding is that infant signal production shows signs of compositionality months before the advent of speech. This conclusion is based on the finding that infants growing up in a rural community in the Peruvian Amazon, combined their gestural points with acoustically distinct classes of vocalisations that discriminated between imperative and declarative intentions. Subsequent experiments demonstrated that these subtle acoustic differences were understood by European adults, demonstrating that the compositional machinery is already operational before speech, the likely evolutionary remnants of our old primate heritage [47]. *PIs: Henderson, Stoll, Townsend; Collaborating PIs: Bangerter, Burkart, Daum, Zuberbühler; Staff: Behjati, Mine, Dickerman*

## 8 Cooperation Project

In the Cooperation Project we seek to better understand the cooperative roots of language through focusing on three interrelated levels: cognition, behaviour and communication.

In **WP MentalStates** we propose a framework based on script theory to understand the goals of agents and the means by which these goals are achieved. This approach helps circumvent a plethora of issues that have plagued the study of the cognitive mechanisms underlying language from over-intellectualisation to inconsistent replication. Critically, results from field experiments with wild apes, and eye-tracking experiments with children and captive apes are generally in line with the predictions of our framework and support the notion that a script-based account of the socio-cognitive foundations of language provides another interesting framework beyond that grounded in mental-state attribution skills. [48]. *PIs: Clément, Glock, Zuberbühler; Collaborating PIs: Bangerter, Burkart, Stoll, Townsend, Van Schaik, Wild; Staff: Taylor, Xhemaili, Guardia, Alexander; Internal Collaborator: Gönül*

**WP JointAction** revisits a dominant proposal in the field of language evolution that the proclivity to engage in shared (joint) actions was a key driver in promoting the evolution of language. Our empirical work in apes and monkeys shows that many of the components of joint action we know to be crucial in human communicative interactions are also at play when primates behaviourally coordinate, and particularly so in cooperative breeding primates [49]–[52]. These findings support the idea that the so-called "interaction engine" which is so crucial to scaffolding human communication likely evolved long-before the emergence of language itself. *PIs: Bangerter, Burkart; Collaborating PI: Stoll; Staff: Morozova, Genty, Adriaense; Internal Collaborators: Miss*

**WP Accommodation**, finally, assessed the process of inter-individual vocal convergence in the service of coordination. The assumption is that vocal accommodation compromises vocal identity, which creates a dilemma of opposing needs, a problem addressed comparatively. First, cooperative and voice recognition tasks showed that in large human groups spectral and temporal voice features became predictably more similar, while speakers continued to signal identity through voice quality features [53]. Second, a method comparison for measuring acoustic variation in mammalian vocal behaviour was conducted [54]. When applied to marmosets, it was found that closeness/individuality trade-offs varied across call types [55], [56]. Finally, in a model comparison it was found that the 'dynamic auditory template matching model' provided the best mechanistic explanation of mutual adjustment in this primate (Figure 11) [57]. *PIs: Burkart, Dellwo; Collaborating PIs: Stoll, Van Schaik; Staff: Pellegrino, Wierucka, Phaniraj; Internal Collaborators: Roswandowitz, Froehlich, Cruz; External Collaborator: Thayabaran Kathiresan (Telepathy)*

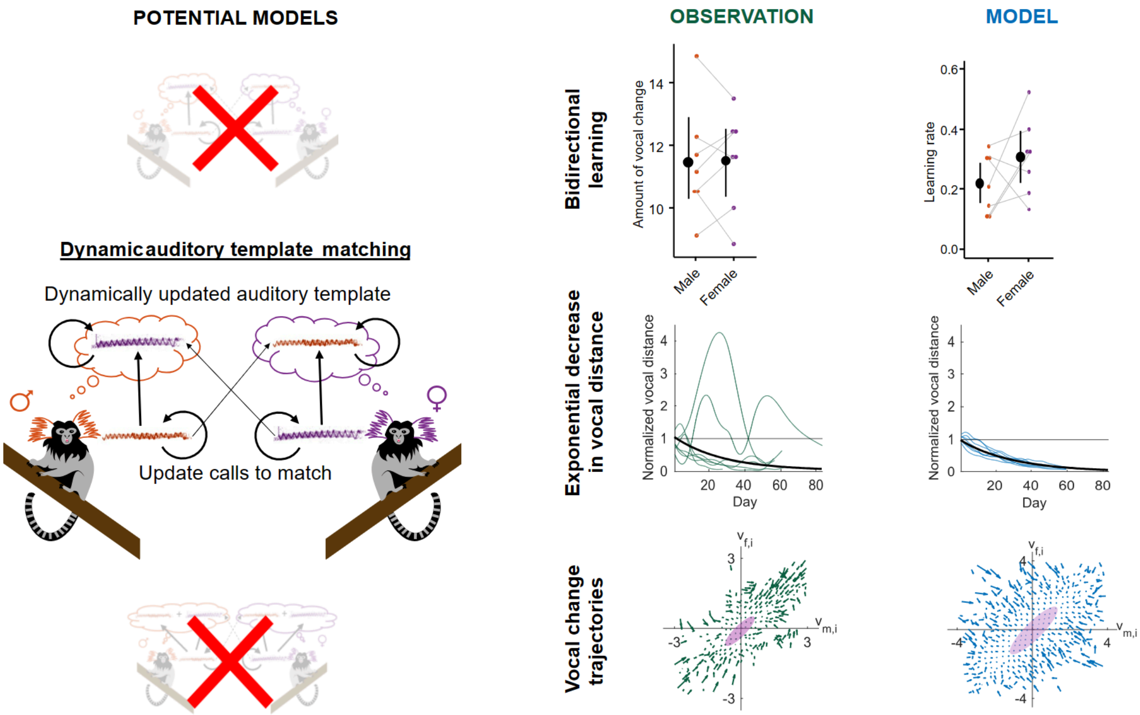



Figure 11. Out of the three candidate models, the dynamic auditory template matching model best explains marmoset vocal accommodation. The model suggests that marmosets keep track of their partner's calls and continuously update their own calls to match them, and displays all the properties of marmoset vocal accommodation patterns: bidirectional learning, an exponential decrease in vocal distance with time and synchronous vocal trajectories.

## 9 Digitization Project

With the Digitization Project we have focused on the future of human language and specifically how our technology may impact linguistic system(s) and their ontogenetic development.

In **WP CompuLang** we have applied retrospective studies, lexico-syntactic analysis, and comparison to human writing to investigate how machine-generated language (i.e machine translation systems, auto-completion tools, and recent language models including ChatGPT) impacts human linguistic intuition and production. Analyses indicate that whilst machine-produced texts still suffer from standardisation and monotonicity in terms of syntax, certain commercial systems produce texts with levels of lexical and morphological richness comparable to those of human translators. When interacting with machines, humans tend to adapt their own language to optimise their interactions. Our studies suggest that the longer the interactions between humans and machines, the stronger the adaptation [58]. Our latest results [59] show that users seem to be willing to compromise on concision if it improves significance, as reflected by greater variation in dependency lengths than found in ChatGPT-generated text. *PIs: Lovis, Volk; Collaborating PIs: Dellwo, Henderson, Laganaro, Stoll; Staff: Göhring, Zaghir, Shaitarova; Internal Collaborators: Goldmann, Gaudet-Blavignac, Bjelogrlic*

In **WP Edugames** we apply prospective studies, leveraging novel digital technologies such as video games [60] and ask how these may be implemented to alter language functions, using literacy acquisition as an example. Through developing a child-friendly video game targeting domain-general skills such as attentional control and executive functions, we show that only tens of hours of play over a few weeks is sufficient to boost reading skills in a large sample of typically developing Italian speaking children (Figure 12). Of note, improvements were seen in standardised surveys of reading just upon completion of the training, but also through school grade performance, albeit with a protracted delay (12 to 18 months after training completion) [61]. This work has led to the development of new reading assessments that integrate Automatic Speech Recognition for easy deployment [62] either over the internet or in a group of children, illustrating the applied potential and critical societal relevance of the work conducted more broadly within the EvolvingLanguage NCCR. *PIs: Bavelier, Zesiger; Collaborating PIs: Brem, Clément, Golestani, Laganaro, Stoll; Staff: Piton, Cohen, Pasqualotto, Röthlisberger, Radonjic; External Collaborator: Irene Altarelli (U. Paris-Cité)*

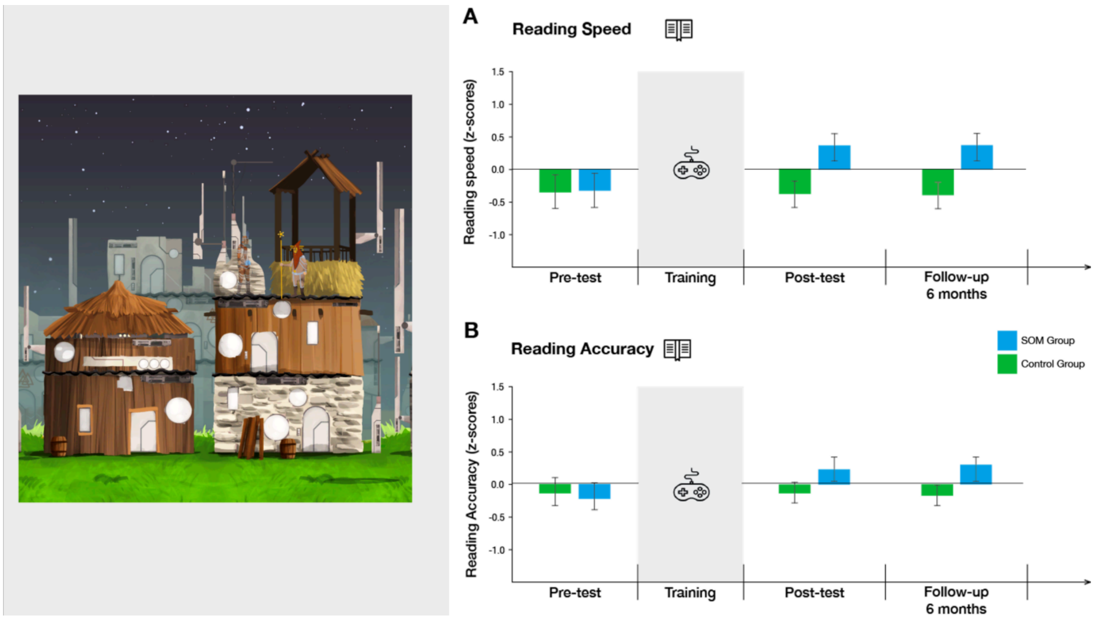



Figure 12. Illustration of the video game used in a 12 hours over 6 weeks intervention study to facilitate reading skills via attentional enhancement (Left). While both groups showed similar performance pre-training, the experimental group (LoH in blue) showed greater improvement in reading speed and accuracy as compared to the control group (in green) after training which was still visible at 6-month follow-up.

## 10 Transversal Task Forces (TTFs)

Over the last three years our TTFs have been serving the methodological, ethical and technological demands of all projects. Key highlights include the following:

Within the **TTF Ethics** (PIs Hurst and Wild) we have built a novel ethics consultation service for researchers and offered its service to researchers in the NCCR. This service has been widely and productively taken up and helped to put ethical best practice at the centre stage of empirical work in the NCCR. **TTF Concepts** (PIs Glock, Weber) has been conducting critical groundwork in the form of methodological and conceptual integration, ensuring fruitful collaborative research that is not hampered by terminological ambiguity and/or conceptual opacity. To this end, six annual workshops and roundtable discussions have been organised, fostering cross-disciplinary discussions and the bottom-up evolution of a shared working vocabulary. As a consequence, the TTF Concepts has played a critical role in a number of theoretical and position papers published within the NCCR including clarifying conceptualisations of arbitrariness in language [63], but also notions of signal-meaning mapping in animal communication [13].

The **TTF Datascience** (PIs Stoll, Hahnloser, Furrer, Merlo, Van De Ville)**,** which serves as the backbone for all empirical work in the NCCR, has supported researchers on topics such as eye-tracking data preprocessing, generative language model implementation, leveraging large language models for low-resource languages, and statistical modelling. WP Databases has worked on a unified approach for managing research data across species, languages, and brain imaging modalities deploying openBIS (https://openbis.ch/) in a pipeline to allow for customizable data management for our NCCR and ensure FAIR (Findable, Accessible, Interoperable, and Reusable) datasets for all the research data generated by the NCCR. Also, WP Databases has been developing and managing large databases. First, we have continued to expand the ACQDIV Database, comprising data of child language acquisition in typologically maximally diverse languages (https://www.acqdiv.uzh.ch/en/resources.html). Second, we have developed a unified metadata schema for the CallBase project, allowing us to aggregate data from various sources and species and for it to be enriched with vocalisations standardly segmented by tools we are currently developing. Since we are also adding song and other vocalisation data we rename the database *VoCallBase*. WP MachineLearning has developed a variational auto-encoder-based platform for exploring distributed text representations, supporting compositionality and disentanglement whilst also creating a deep learning model for animal vocal segmentation based on the Whisper model [64], which has already been successfully used for segmenting zebra finch syllables. Lastly, WP Statistics has developed novel statistical frameworks in support of ten different WPs, for example modelling eye-tracking data with measurement error [8], assessing the impact of sample size on child speech statistics (manuscript in preparation by Ringen, Stoll and Furrer) or reducing linguistic typology data in order to minimize NAs while maximising phylogenetic and geographical diversity (R package "Densify" by Lischka, Graff, Bickel and Furrer, near completion). Additionally, we are committed to capacity-building within the NCCR, offering a "Statistics and Causal Inference" reading group, all the meetings of which were recorded and shared.

As part of the **TTF Technology** (PIs Magimai-Doss, Bavelier, Brem, Hahnloser) the WP Automatic Speech Recognition has made key progress in developing speech and audio processing methods/technologies to aid in the study of human speech communication and animal communication. These developments have, for example, been integrated into the WP EduGame to assess child reading skills and provide a tool for pronunciation assessment at the phonetic level [62] and the WP Accommodation to aid in the automated classification of signaler identity in marmoset monkeys [65]. The WP NeuroFeedback has also been critical in advancing closed-loop brain-computer interface systems for the decoding of speech-related brain signals through the application of adaptive, as opposed to static, classifier methods [66].

## New Developments in Research Structure: Special Interest Groups (SIGs)

Over the last 3 years a number of Special Interest Groups (SIGs) have been founded, targeting research questions relevant to the NCCR agenda but not (yet) currently addressed within any of the ongoing work packages.

The **SIG Cross-linguistic Studies of Reading and Reading-related Disorders** (Bavelier, Brem, Zesiger, Giraud) explores how reading might have evolved from domain-general or language-specific mechanisms and tries to unpack the many factors contributing to individual differences in reading ability across languages of different orthographic transparency. Over the last 12 months and with the help of an SNF Scientific Exchange grant the SIG organised a workshop on "Cross-linguistic reading interventions" in Fall 2022 in Geneva. The workshop achieved its objective of identifying the fundamental elements of the joint reading training study design and setting the stage for reaching an agreement on writing a pre-registration article for the study. This workshop also served to establish an international network of collaborators, including Prof. Bavelier (UNIGE), Prof. Brem (UZH), Prof. Zoccolotti (La Sapienza U. of Rome; IT), Prof. Ziegler (Aix-Marseille U.; FR) and Prof. Altarelli (Paris Cité U.; FR). Concrete next steps include pre-registering the planned intervention study to be carried out in parallel in Italian and in French, the most and least orthographically transparent languages used by a majority of children in our educational system.

The focus of the **SIG Metaphors** (Stoll, Widmer, and Narayanan from Google Research Zurich) is on how meaning is extended via metaphorization and colexification in ontogeny and (linguistic) phylogeny. Both metaphors and colexification allow for the expansion of meaning without introducing new forms (vocabulary). This makes them powerful tools for creativity without additional formal affordances. The SIG Metaphors comprises two inter-linked sub-projects: Subproject 1 focuses on the role of metaphors in ontogeny, subproject 2 on colexification in IndoEuropean, a suitable language family testbed with a large database of well documented languages. To realise these projects, the SIG Metaphors has independently secured external funding from the Gerda Zeltner Fund at UZH for 2 PhD students. In the ontogeny project a paper is currently being finalised on the role of metaphors in three genres: child-directed speech, spoken language and written language. Through applying Melbert [67] and concreteness algorithms to three large naturalistic corpora, child-directed speech was shown to have significantly less metaphoric but more concrete content than the other two genres. This shows that child-directed speech is indeed simpler on a semantic level than other linguistic genres. In the phylogeny project, approx. 250 colexifications that are found in more than ten languages were identified and enriched with cognacy data. Phylogenetic models are currently being fitted in order to determine the evolutionary flexibility of colexified concepts and their interaction with morphological material and cultural contexts.

A related but more conceptually-oriented (than empirically) SIG, the **SIG Meaning** (Burkart, Glock) tackles the concept of meaning from a cross-species comparative perspective, addressing questions on the nature of vocal or gestural units involved in meaning-differentiation in addition to the role of pragmatic competencies and theory of mind in the non-human animal capacity to attribute meaning. An interdisciplinary manuscript is currently being written up reviewing key research related to these questions, proposing a novel approach for the investigation of meaning within the field of language evolution studies.

The **SIG Cetacean Communication** (Zuberbühler, Hahnloser) aims to expand the NCCR's comparative research to mammals that transitioned back to the aquatic realm around 50 million years ago. These animals developed a sophisticated communication system in a very different acoustic environment with 3D wave fields playing a pivotal role. The SIG's focus has been two-fold: firstly, developing technologies to record communicative signals and compiling datasets for research using emerging data science methods and secondly, initiating global collaborations with scientists for the upcoming initiatives. To facilitate this work the SIG Cetacean Communication has also received two Innovation Grants to improve array recording technology and signal processing algorithms through student projects and to fund an expedition to gather a substantial dataset dedicated to exploring signal information capacity and dialect dynamics.

The **SIG Cultures and Action Structures** (Bickel, Giraud, Meyer, Grandjean) has served as a think tank to reshape our research agenda in syntax, discussing new ways to approach classical comparisons between language, tool production and action planning. The results are the description of the Phase 2 Syntax project.

Finally, the **SIG Mechanisms of Vocal Learning in Ovo and Utero** (Hahnloser, Hervais-Adelman, Townsend) focuses on the early ontogeny of vocal communication, supported by an innovation grant to investigate the combinatorial dynamics of infant marmoset vocalisations. In humans, a rapid increase in melodic complexity of cries is observed in the first weeks after birth, consistent with the possibility that cries are not merely manifestations of straightforward alarm calls, but rather carry the trace of an incipient combinatorial potential of prosodic building blocks [68]. Comparing the developmental trajectory of marmoset vocalisations with those of infant humans will provide insight into whether the development of combinatorial complexity is a precocious hallmark of a combinatorial system (as present in humans) or can be assigned to the development of increasingly sophisticated vocal motor control (as present in both humans and marmosets). After a presubmission inquiry the SIG team was encouraged to prepare an opinion piece highlighting the neglected role of in-utero/in-ovo sound exposure on vocal development for PLoS Biology.

## National, International and within NCCR Collaborations

Given the transdisciplinary nature of our NCCR, from the very beginning we have nurtured and fostered collaborations on all levels. At the national level, the NCCR supports the strategic partnership between UZH and UNIGE by fostering teams of PIs that work together between the two institutions and that use each other's resources, for example intracranial sEEG research at the hospital in Geneva co-involving PIs from Zurich; phylogenetic modelling work in Zurich co-involving PIs from Geneva; or joint EEG research in Vietnam using resources from both the HNP in Geneva and the LiRI in Zurich. Beyond the two home institutions, the NCCR has deepened collaborations with several other Swiss institutions, for example, between U Neuchâtel and UZH on primate cognition or between U Basel and UZH on philosophy of animal minds. The NCCR has also made entirely new lines of collaborative research in Switzerland possible, for example between the Institute for Multilingualism in Fribourg and neuroscience work on language aptitude in Geneva or between the IDIAP Institute in Martigny and work on language acquisition in Zurich on tracking joint attention. Collaboration also involves partners outside academia. We have close collaboration with Basel Zoo where we have set up and now use facilities to administer eye-tracking and touch screen experiments with great apes. With the Swiss Society for Linguistics we work on developing materials for promoting language science at high schools (*Gymnasium, gymnase*) (Chapter 5.1.). For industrial partners, see our report of Knowledge and Technology Transfer (Chapter 5.2)

As attested by our publications, all our PIs also maintain extensive international collaborations. Centres with more long-term collaboration, involving several PIs, include MPI Nijmegen, MPI-EVA Leipzig, ICREA Barcelona, MIT, U Vienna, Harvard U, U Heidelberg, U Tübingen, UCLA, Tokai U Japan, or the CHU Hospital in Bordeaux, to name a few. Of particular importance are our partnerships with institutions where we conduct fieldwork, fostering inclusivity in our research through joint research and teaching: Makerere University in Uganda, University of Cocody in Ivory Coast, the Vietnamese National University in Hanoi, and the Pontificia Universidad Católica del Perú in Lima.

Within the NCCR, collaboration chiefly happens within specific WPs. From the start of the NCCR we have followed the policy of exclusively funding teams of PIs, mostly from different disciplines and often also from different universities. However, project meetings, retreats and summer schools repeatedly inspired collaborations *between* work packages, as documented by joint publications (Figure 3). An additional forum for cross-WP collaboration are the Special Interest Groups (SIGs) that we have launched over the years (see Section 4.3).

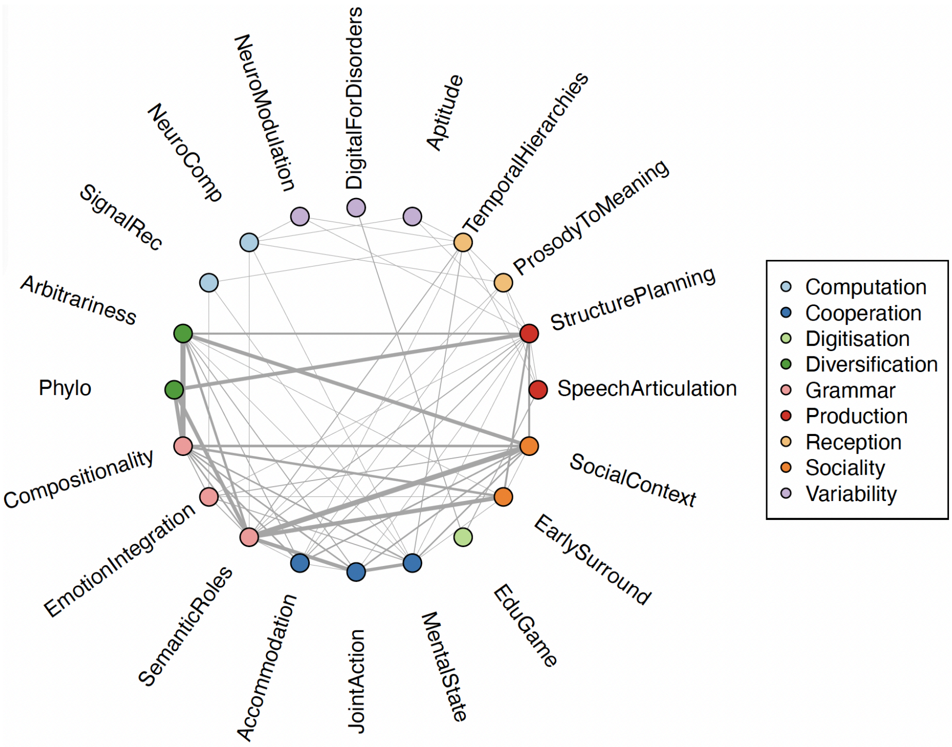



Figure 3. Additional collaboration between WPs as evidenced by joint publications. This type of collaboration developed spontaneously from NCCR-wide meetings and retreats.

While these data reveal the sheer scale of interactions, they say little about the intellectual undercurrent and cross-fertilisation attached to each collaboration. Examples are numerous — computational scientists finding common language with neuroscientists when trying to compare silicon and biological systems and to model neural processes (WP Neurocomp); linguists, inspired by findings in animal cognition, asking new questions about the lower bounds of compositionality (WP Compositionality); primatologists, inspired by findings in linguistics, testing new hypotheses about great ape agency perception (WP SemanticRoles); social psychologists and animal behaviour researchers adopting shared methods and concepts (WP MentalState), language acquisition researchers venturing deeper into prelinguistic communication, with conceptual tools from animal communication research (WP EarlySurround); neuroscientists rethinking the foundations of theta oscillations through phylogenetic modelling of communication rhythms (WP TemporalHierarchies). These are encouraging advances and we foresee to fully reap the intellectual benefits of our Phase 1 efforts into Phase 2.

# References


[1] Bosshard AB, Burkart JM, Merlo P, Bickel B, Townsend SW. Submitted. Beyond bigrams: Call sequencing in the common marmoset (Callithrix Jacchus) vocal system. *Royal Society Open Science*.

[2] Merlo P, An A, Rodriguez MA. 2022. Blackbird's language matrices (BLMs): a new benchmark to investigate disentangled generalisation in neural networks. *arXiv*. doi:10.48550/arXiv.2205.10866.

[3] Egurtzegi AM, Sauppe S, Isasi-Isasmendi A, Martinez de la Hidalga G, Bornkessel-Schlesewsky I, Schlesewsky M, Laka I, Meyer M, Bickel B, Andrews C. under re-review. The effect of animacy on the Agent Preference: self-paced reading evidence from Basque. *Memory & Cognition*.

[4] Isasi-Isasmendi A, Andrews C, Flecken M, Laka I, Daum MM, Meyer M, Bickel B. 2023. The Agent Preference in Visual Event Apprehension. *Open Mind*.

[5] Huber E, Sauppe S, Isasi-Isasmendi A, Bornkessel-Schlesewsky I, Merlo P, Bickel B. 2023. Surprisal from language models can predict ERPs in processing predicate-argument structures only if enriched by an agent preference principle. *Neurobiology of Language*.

[6] Sauppe S, Næss Å, Roversi G, Meyer M, Bornkessel-Schlesewsky I, Bickel B. under re-review. An Agent-First Preference in a Patient-First Language during Sentence Comprehension. *Cognitive Science*.

[7] Brocard S, Wilson VAD, Chloé B, Zuberbühler K, Bickel B. Under review. Foundations of language processing: A hominid preference for animate agency. *PNAS*.

[8] Wilson VAD, Sauppe S, Brocard S, Ringen EJ, Daum MM, Wermelinger S, Andrews C, Isasi-Isasmendi A, Bickel B, Zuberbühler K. Under review. Primate Origins of Human Event Cognition.

[9] Wilson VAD, Zuberbühler K, Bickel B. 2022. The evolutionary origins of syntax: Event cognition in nonhuman primates. *Sci. Adv.* 8(25):eabn8464. doi:10.1126/sciadv.abn8464.

[10] Wilson VAD, Sauppe S, Brocard S, Ringen E, Daum MM, Wermelinger S, Andrews C, Isasi-Isasmendi A, Bickel B, Zuberbühler K. 2023. Primate Origins of Human Event Cognition.

[11] Sauppe S, Næss Å, Roversi G, Meyer M, Bornkessel-Schlesewsky I, Bickel B. 2023. An Agent-First Preference in a Patient-First Language During Sentence Comprehension. *Cognitive Science*.

[12] Huber E, Sauppe S, Isasi-Isasmendi A, Bornkessel-Schlesewsky I, Merlo P, Bickel B. 2023. Surprisal from Language Models Can Predict ERPs in Processing Predicate-Argument Structures Only If Enriched by an Agent Preference Principle. *Neurobiology of Language*.

[13] Watson SK, Filippi P, Gasparri L, Falk N, Tamer N, Widmer P, Manser M, Glock H-J. 2022. Optionality in animal communication: a novel framework for examining the evolution of arbitrariness. *Biological Reviews*. 97(6):2057–75. doi:10.1111/brv.12882.

[14] Ranacher P, Neureiter N, van Gijn R, Sonnenhauser B, Escher A, Weibel R, Muysken P, Bickel B. 2021. Contact-tracing in cultural evolution: a Bayesian mixture model to detect geographic areas of language contact. *J. R. Soc. Interface*. 18(181):20201031. doi:10.1098/rsif.2020.1031.

[15] Neureiter N, Ranacher P, Efrat-Kowalsky N, Kaiping GA, Weibel R, Widmer P, Bouckaert RR. 2022. Detecting contact in language trees: a Bayesian phylogenetic model with horizontal transfer. *Humanities and Social Sciences Communications*. 9(1):1–14. doi:10.1057/s41599-022-01211-7.

[16] Cathcart C, Herce B, Bickel B. 2022. Decoupling Speed of Change and Long-Term Preference in Language Evolution: Insights From Romance Verb Stem Alternations. *Proceedings of the Joint Conference on Language Evolution (JCoLE)*. doi:10.17617/2.3398549.

[17] Cathcart C, Bickel B, Raviv L, Boeckx C. 2024. Linguistic evolution in time and space. In *The Oxford Handbook of Approaches to Language Evolution*. Oxford: Oxford University Press.

[18] Barbieri C, Blasi DE, Forkel R, Hammarström H, Wichmann S, Greenhill SJ, Bickel B, Gray RD. 2018. *GELATO: Genes and Languages Together*. Work. Pap.

[19] Cathcart C. 2023. Rate variation in language change: Toward distributional phylogenetic modeling. In *Biocultural Implications: An Agenda for Integrative Approaches*, ed Karakostis, Alexandros, G Jäger. Tübingen: Kerns Verlag.

[20] Barbieri C, Blasi D, Forkel R. 2022. GEnes and LAnguages TOgether. doi:10.5281/ZENODO.7233267.

[21] Barbieri C, Blasi DE, Arango-Isaza E, Sotiropoulos AG, Hammarström H, Wichmann S, Greenhill SJ, Gray RD, Forkel R, Bickel B, Shimizu KK. 2022. A global analysis of matches and mismatches between human genetic and linguistic histories. *Proceedings of the National Academy of Sciences*. 119(47):e2122084119. doi:10.1073/pnas.2122084119.
[22] Efrat-Kowalsky N, Ranacher P, Neureiter N, Widmer P, Bickel B. Under Review. Oldest Attested Languages in the Near East Reveal Deep Transformations in Linguistic Landscapes.
[23] Hovsepyan S, Olasagasti I, Giraud A-L. 2020. Combining predictive coding and neural oscillations enables online syllable recognition in natural speech. *Nature Communications*. 11(1):3117. doi:10.1038/s41467-020-16956-5.
[24] Hovsepyan S, Olasagasti I, Giraud A-L. 2022. Rhythmic modulation of prediction errors: a possible role for the beta-range in speech processing. *bioRxiv*. doi:10.1101/2022.03.28.486037.
[25] Su Y, MacGregor LJ, Olasagasti I, Giraud A-L. 2023. A deep hierarchy of predictions enables online meaning extraction in a computational model of human speech comprehension. *PLoS Biol*. 21(3):e3002046. doi:10.1371/journal.pbio.3002046.
[26] Henderson J, Fehr FJ. 2023. A VAE for Transformers with Nonparametric Variational Information Bottleneck. (available at https://openreview.net/forum?id=6QkjC_cs03X).
[27] Piette T, Barbieri C, Grandjean D, Déaux ÉC, Giraud AL. 2022. Theta rhythm is widespread in vocal production across the animal realm. *Proceedings of the Joint Conference on Language Evolution*.
[28] Déaux ÉC, Piette T, Gaunet F, Legou T, Arnal L, Giraud A-L. submitted. Dogs' sensory-motor tuning shapes dog-human vocal interactions.
[29] Oderbolz C, Sauppe S, Meyer M. Neural tracking of the prosodic hierarchy: Evidence from the comprehension of rhythmically regular and irregular sentences.
[30] Degano G, Donhauser P, Gwilliams L, Merlo P, Golestani N. 2023. Speech prosody enhances the neural processing of syntax. *BioRxiv*. doi:10.1101/2023.07.03.547482.
[31] Egurtzegi A, Blasi DE, Bornkessel-Schlesewsky I, Laka I, Meyer M, Bickel B, Sauppe S. 2022. Cross-linguistic differences in case marking shape neural power dynamics and gaze behavior during sentence planning. *Brain and Language*. 230:105127. doi:10.1016/j.bandl.2022.105127.
[32] Sauppe S, Choudhary KK, Giroud N, Blasi DE, Norcliffe E, Bhattamishra S, Gulati M, Egurtzegi A, Bornkessel-Schlesewsky I, Meyer M, Bickel B. 2021. Neural signatures of syntactic variation in speech planning. *PLOS Biology*. 19(1):e3001038. doi:10.1371/journal.pbio.3001038.
[33] Morucci P, Nabé M, Sauppe S, Meyer M, Proix T, Bickel B, Giraud A-L. Under review. Neural Encoding of Syntactic Structures during Natural Speech Planning and Production. *Society for the Neurobiology of Language*.
[34] Lancheros M, Friedrichs D, Laganaro M. 2023. What Do Differences between Alternating and Sequential Diadochokinetic Tasks Tell Us about the Development of Oromotor Skills? An Insight from Childhood to Adulthood. *Brain Sciences*. 13(4):655. doi:10.3390/brainsci13040655.
[35] Lancheros M, Atanasova T, Laganaro M. 2023. When does speech planning rely on motor routines? ERP comparison of speech and non-speech from childhood to adulthood. *Language, Cognition and Neuroscience*.
[36] Friedrichs D, Dellwo V. 2023. Speaking fast and slow: Evidence for anatomical influence on temporal properties of speech. *The Journal of the Acoustical Society of America*.
[37] Kepinska O, Caballero J, Oliver M, Marks RA, Haft SL, Zekelman L, Kovelman I, Uchikoshi Y, Hoeft F. 2023. Language combinations of multilinguals are reflected in their first-language knowledge and processing. *Scientific reports*. 13(1):1947. doi:10.1038/s41598-023-27952-2.
[38] Kepinska O, Dalboni Da Rocha J, Tuerk C, Hervais-Adelman A, Price CJ, Green D, Golestani, Narly. 2022. Auditory cortex accommodates multilingualism. *14th Annual Meeting of the Society for the Neurobiology of Language*.
[39] Farcy C, Chauvigné L, Laganaro M, Corre M, Ptak R, Guggisberg AG. Submitted. Neural mechanisms underlying improved new-word learning with high-density transcranial direct current stimulation.
[40] Haugg A, Frei N, Menghini M, Stutz F, Steinegger S, Röthlisberger M, Brem S. 2023. Self-regulation of visual word form area activation with real-time fMRI neurofeedback. *Sci Rep*. 13(1):9195. doi:10.1038/s41598-023-35932-9.
[41] Franco J, De Deckersberg M, Bourdon V, Laganaro M. Submitted. Word learning and consolidation through immersive virtual reality.
[42] Fraga-González G, Willinger D, Gehrig V, Fri N, Brem S. Submitted. Neural representation of association strength and prediction error during novel symbol-speech sounds learning.
[43] Schick J, Fryns C, Wegdell F, Laporte M, Zuberbühler K, Van Schaik CP, Townsend SW, Stoll S.

2022. The function and evolution of child-directed communication. *PLOS Biology*. 20(5):e3001630. doi:10.1371/journal.pbio.3001630.
[44] Dickerman M, Mine J, Behjati M, Henderson J, Townsend SW, Stoll S. Submitted. Pragmatic competence in language evolution.
[45] Mine JG, Wilke C, Zuberti C, Bejhati M, Machanda Z, Dickerman M, Manser A, Stoll S, Slocombe KE, Townsend SW. 2023. Vocal-visual combinations in wild chimpanzees. *bioRxiv*. doi:10.1101/2023.05.24.541909.
[46] Behjati M, Henderson J. 2021. Inducing Meaningful Units from Character Sequences with Slot Attention. *arXiv:2102.01223 [cs]*. (available at http://arxiv.org/abs/2102.01223).
[47] Schick J, Saldana C, Kathiresan T, Zuberbühler K, Stoll S. under review. Prelinguistic Compositionality: Multimodal Communication in Infants as a Developmental Precursor to Language Acquisition. *Science Advances*.
[48] Taylor D, Gönül G, Alexander C, Clément F, Zuberbühler K, Glock HJ. Reading Minds Or Reading Scripts? De-intellectualising Theory of Mind. *Biological Reviews*. under Reveiw:
[49] Genty E. in prep. Gesture amplitudes reflect audience design in ape gestural communication.
[50] Adriaense JEC, Burkart JM. in prep. Parental care as joint action in common marmosets: coordination during infant transfers.
[51] Adriaense JEC, Burkart JM. in prep. Social play as joint action in common marmosets: exploring indicators of joint commitment with an interaction framework approach.
[52] Miss FM, Meunier H, Burkart JM. 2022. Primate origins of corepresentation and cooperative flexibility: A comparative study with common marmosets (Callithrix jacchus), brown capuchins (Sapajus apella), and Tonkean macaques (Macaca tonkeana). *Journal of Comparative Psychology*. 136(2):199–212. doi:10.1037/com0000315.
[53] Pellegrino E, Dellwo V. Accepted Speakers are more cooperative and less individual when interacting in larger group sizes. *Frontiers in Psychology*. doi:doi: 10.3389/fpsyg.2023.1145572.
[54] Wierucka K, Murphy D, Watson S, Falk N, Fichtel C, León J, Leu ST, Kappeler P, Briefer EF, Manser M, Phaniraj N, Scheumann M, Burkart JM. Same data, different results? A comparative evaluation of automated analysis methods for bioacoustics research.
[55] Zürcher Y, Willems EP, Burkart JM. 2021. Trade-offs between vocal accommodation and individual recognisability in common marmoset vocalizations. *Sci Rep*. 11:15683. doi:10.1038/s41598-021-95101-8.
[56] Phaniraj N. 2022. *Machine learning approaches for tracking changes in marmoset vocalizations and social behaviors*. Thesis thesis. (available at http://dr.iiserpune.ac.in:8080/xmlui/handle/123456789/6903).
[57] Phaniraj N, Wierucka K, Burkart JM. in prep. The temporal dynamics of marmoset vocal learning suggests a dynamic template matching mechanism.
[58] Zaghir J, Goldman J-P, Bjelogrlic M, Gaudet-Blavignac C, Lovis C. 2022. Caregivers Interactions with Clinical Autocomplete Tool: A Retrospective Study. *Studies in Health Technology and Informatics*. 295:132–35.
[59] Zaghir J, Bjelogrlic M, Goldman J-P, Ehrsam J, Gaudet-Blavignac C, Lovis C. in progress. Language adaptation and Zipf's Principle of Least Effort: accuracy against efficiency.
[60] Pasqualotto A, Parong J, Green CS, Bavelier D. 2022. Video Game Design for Learning to Learn. *International Journal of Human–Computer Interaction*, pp. 1–18. doi:10.1080/10447318.2022.2110684.
[61] Pasqualotto A, Altarelli I, De Angeli A, Menestrina Z, Bavelier D, Venuti P. 2022. Enhancing reading skills through a video game mixing action mechanics and cognitive training. *Nat Hum Behav*. 6(4):545–54. doi:10.1038/s41562-021-01254-x.
[62] Piton T, Hermann E, Pasqualotto A, Cohen M, Magimai M, Bavelier D. 2023. Using Commercial ASR Solutions to Assess Reading Skills in Children: A Case Report.
[63] Gasparri L, Filippi P, Wild M, Glock H-J. 2022. Notions of arbitrariness. *Mind & Language*. doi:10.1111/mila.12443.
[64] Radford A, Kim JW, Xu T, Brockman G, Mcleavey C, Sutskever I. 2023. Robust Speech Recognition via Large-Scale Weak Supervision. *Proceedings of the 40th International Conference on Machine Learning*, pp. 28492–518.
[65] Phaniraj N, Wierucka K, Zürcher Y, Burkart JM. 2022. Optimising source identification from marmoset vocalisations with hierarchical machine learning classifiers. *bioRxiv*. 2022.11.19.517179. doi:10.1101/2022.11.19.517179.
[66] Wu S, Kinkini B, Giraud A-L, Marchesotti S. Submitted. Adaptive LDA Classifiers improve syllables decoding during brain-computer interface control.
[67] Choi M, Lee S, Choi E, Park H, Lee J, Lee D, Lee J. 2021. MelBERT: Metaphor Detection via

Contextualized Late Interaction using Metaphorical Identification Theories. *arXiv*. doi:10.48550/arXiv.2104.13615.
[68] Wermke K, Robb MP, Schluter PJ. 2021. Melody complexity of infants' cry and non-cry vocalisations increases across the first six months. *Sci Rep*. 11(1):4137. doi:10.1038/s41598-021-83564-8.

# Annex 1: Status of Structural Measures Implementation

Table 21. Status of Structural Measures in July 2023

| Planned measures according to annex 3 of the NCCR contract for phase I | Current status of implementations and comments |
|---|---|
| **University of Zurich** | |
| *Faculty* | |
| Open rank position Digital Linguistics (in-kind) | completed (position filled with Lena Jäger in the rank of an associate professor in August 2020) |
| Assistant Professor (TT) Descriptive and Anthropological Linguistics (cash) | completed (position filled with John Mansfield in the rank of an assistant professor in August 2023) |
| Associate Professor Neuroscience of Language (in-kind) | completed (position filled with Martin Meyer in September 2021) |
| Assistant Professor (TT) Genetics of Language (cash) | in progress (call goes out in Fall 2023) |
| Associate Professor Primate Communication (cash) | completed (position filled with Simon Townsend in the rank of an associate professor planed in October 2023) |
| Open rank position Paleoanthropology (succession Zollikofer) (in-kind) | in progress (call goes out in Fall 2023) |
| *Infrastructure* | |
| Technical platform Linguistic Research Infrastructure (LIRI) (cash+in-kin) | completed (all data science positions have been filled; TTF has been organisationally integrated into LiRI under the label “NCCR@LiRI”; in-kind contributions are being used to expand technology offered by LiRI) |
| **University of Geneva** | |
| *Faculty* | |
| Associate Professor Language Sciences (for former SNSF professor) | completed (position had been filled with Narly Golestani, who left U Geneva in March 2021; position has now been filled again with Valentina Borghesani in September 2022) |
| Assistant Professor (non-TT) Neural Data Science (in-kind and cash) | completed (position filled with Alexis Hervais-Adelman who will start in September 2023) |
| Infrastructure | |

| Planned measures according to annex 3 of the NCCR contract for phase I | Current status of implementations and comments |
|---|---|
| MEG platform | completed (MEG is fully operational since beginning 2023) |
| Data Science platform | in progress (new technological are being developed and implemented) |
| *Additional Measures* | |
| PhD School Biology of Language | completed (catalogue with relevant courses from participating institutions has been created) |

# Annex 2: WPs collaborations

Table 22. List of collaborations between our WPs as evidenced by joint publications. Figure 3 is based on this list.

| WP | collaborating WPs |
|---|---|
| Accommodation | ProsodyToMeaning, StructurePlanning, SocialContext, JointAction, SemanticRoles, Arbitrariness, TemporalHierarchies, Compositionality |
| Aptitude | TemporalHierarchies, StructurePlanning, NeuroComp |
| Arbitrariness | StructurePlanning, SocialContext, Accommodation, JointAction, SemanticRoles, MentalState, EarlySurround, Compositionality |
| Compositionality | SocialContext, JointAction, Accommodation, SemanticRoles, MentalState, EarlySurround, Arbitrariness, Phylo, EmotionIntegration |
| DigitalForDisorders | EduGame |
| EarlySurround | StructurePlanning, SocialContext, SemanticRoles, MentalState, Arbitrariness, EmotionIntegration, Compositionality |
| EduGame | SpeechArticulation, DigitalForDisorders |
| EmotionIntegration | StructurePlanning, SocialContext, SemanticRoles, EarlySurround, MentalState, SignalRec, Compositionality |
| JointAction | StructurePlanning, SocialContext, Accommodation, SemanticRoles, MentalState, Arbitrariness, Compositionality |
| MentalState | StructurePlanning, NeuroComp, SocialContext, JointAction, EarlySurround, Arbitrariness, TemporalHierarchies, SignalRec, EmotionIntegration, Compositionality |
| NeuroComp | ProsodyToMeaning, NeuroModulation, SemanticRoles, MentalState, Aptitude |
| NeuroModulation | StructurePlanning, TemporalHierarchies, NeuroComp |
| Phylo | SemanticRoles, StructurePlanning, Compositionality |
| ProsodyToMeaning | StructurePlanning, NeuroComp, Accommodation, SpeechArticulation, SemanticRoles, TemporalHierarchies |
| SemanticRoles | ProsodyToMeaning, StructurePlanning, NeuroComp, SocialContext, Accommodation, JointAction, EarlySurround, Arbitrariness, TemporalHierarchies, Phylo, EmotionIntegration, Compositionality |
| SignalRec | TemporalHierarchies, MentalState, EmotionIntegration |
| SocialContext | StructurePlanning, Accommodation, JointAction, SemanticRoles, EarlySurround, MentalState, Arbitrariness, EmotionIntegration, Compositionality |
| SpeechArticulation | ProsodyToMeaning, TemporalHierarchies, StructurePlanning, EduGame |
| StructurePlanning | ProsodyToMeaning, NeuroModulation, SocialContext, Accommodation, JointAction, SpeechArticulation, SemanticRoles, EarlySurround, Aptitude, MentalState, Arbitrariness, Phylo, EmotionIntegration |
| TemporalHierarchies | ProsodyToMeaning, NeuroModulation, Accommodation, SpeechArticulation, SemanticRoles, MentalState, Aptitude, SignalRec |