Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Special Issue: Annals *Reports*

Original Article

Intermodality differences in statistical learning: phylogenetic and ontogenetic influences

Leona Polyanskaya,[1] Héctor M. Manrique,[1] Arthur G. Samuel,[2,3] Antonio Marín,[4] Azucena García-Palacios,[5,6] and Mikhail Ordin[7]

[1]Departamento de Psicología y Sociología, Universidad de Zaragoza, Teruel, Spain. [2]Department of Psychology, Stony Brook University, New York City, New York. [3]Basque Centre on Cognition, Brain and Language, San Sebastian, Spain. [4]Valencian International University, Valencia, Spain. [5]Department of Basic Psychology, Clinical and Psychobiology, Jaume I University, Castellon, Spain. [6]CIBER Fisiopatología, Obesidad y Nutrición (CIBERObn), Instituto Carlos III, Madrid, Spain. [7]Universität Konstanz, Allgemeine Sprachwissenschaft, Konstanz, Germany

Address for correspondence: Mikhail Ordin, Universität Konstanz, Allgemeine Sprachwissenschaft, Universitätsstraβe 10, 78464 Konstanz, Deutschland. mikhail.ordin@uni-konstanz.de

In Basque–Spanish bilinguals, statistical learning (SL) in the visual modality was more efficient on nonlinguistic than linguistic input; in the auditory modality, we found the reverse pattern of results. We hypothesize that SL was shaped for processing nonlinguistic environmental stimuli and only later, as the language faculty emerged, recycled for speech processing. This led to further adaptive changes in the neurocognitive mechanisms underlying speech processing, including SL. By contrast, as a recent cultural innovation, written language has not yet led to adaptations. The current study investigated whether such phylogenetic influences on SL can be modulated by ontogenetic influences on a shorter timescale, over the course of individual development. We explored how SL is modulated by the ambient linguistic environment. We found that SL in the auditory modality can be further modulated by exposure to a bilingual environment, in which speakers need to process a wider range of diverse speech cues. This effect was observed only on linguistic, not nonlinguistic, material. We conclude that ontogenetic factors modulate the efficiency of already existing SL ability, honing it for specific types of input, by providing new targets for selection via exposure to different cues in the sensory input.

Keywords: statistical learning; redeployment; adaptation; ontogenetic influences; phylogenetic influences; positional memory; transitional probabilities; nature-nurture; modality-specificity; domain-specificity

Introduction

Statistical learning (SL) is a set of neurocognitive mechanisms underlying the ability to extract regularities from the environment. These regularities include recurrent patterns and sequences as well as transitional probabilities (TPs), the probability that one event predicts a subsequent event. SL mechanisms have frequently been explored in the context of language acquisition and processing.[1] Indeed, the neurocognitive mechanisms underlying SL are engaged when humans listen to natural speech, and performance on SL tasks is correlated with linguistic abilities.[2–11] However, SL also operates on nonlinguistic material[12–15] and has been observed in a range of taxonomically different species that do not have a language faculty.[16–19] In sum, SL mechanisms are evolutionarily ancient, making it highly unlikely that they evolved specifically to process linguistic input.

SL allows us to efficiently process environmental stimuli and detect underlying structure.[20–25] These neurocognitive mechanisms emerged in the environment of evolutionary adaptedness (EAA, i.e., the environment, in which neurocognitive mechanisms formed under long-term selection

doi: 10.1111/nyas.14749

pressures). More specifically, SL allows for rapid detection of transitions between longer-lasting steady states in our natural environment.[26,27] Such transitions, characterized by breaches in statistical structure, are likely to require adaptive behavioral responses. This adaptive cycle—in which detecting statistical violations between elements in sequences of events[28] affords optimal behavioral responses—has also honed the evolution of SL mechanisms, which were later redeployed for speech processing. In natural languages, TPs between syllables are often reset at the boundaries between linguistic constituents—words, phrases, and sentences.

If SL mechanisms were shaped in the absence of any language faculty in the EAA, their efficiency should be higher on nonlinguistic than linguistic material. We explicitly tested this hypothesis[29] in a group of Basque–Spanish bilinguals and found that SL was more efficient on speech-like content (sequences of syllables) than nonspeech content (sequences of environmental sounds), although both were based on the same set of statistical regularities. In the visual modality, we found the reverse pattern: the same group showed more efficient SL on nonlinguistic material (fractals) than linguistic material (written syllables), both presented one-by-one in the middle of the screen. This pattern suggests that the faculty of vocal speech may have been relevant for individual fitness in the genus *Homo* for sufficiently long time at a phylogenetic timescale in order to enable adaptive use of SL for the linguistic (speech-like) material in the auditory modality. By contrast, written language, that is, the visual linguistic modality, is a more recent cultural invention, which does not reflect adaptive changes to SL. It appears that signed languages and gestures have not influenced SL in the visual modality, presumably because the speech faculty is predominantly vocal, and nondeaf populations have little contact with signed language.

Speech both shapes and makes up an integral part of the human social environmental niche. Its influence can be studied on the *phylogenetic* timescale of our species as well as on the *ontogenetic* timescale of individual development. In the current study, we aimed to explore ontogenetic influences on the efficiency of SL. We set up an artificial language-learning experiment (similar to our previous experiment, reported above[29]) to test three populations—Basque–Spanish bilinguals, Catalan–Spanish bilinguals, and Spanish monolinguals—in the visual and auditory modalities on both linguistic and nonlinguistic materials. Basque (the Gipuzcoa dialect, which forms the basis for *Batua*, the Standard Basque variety) and Northern Castilian Spanish (spoken in the Basque Country) exhibit phonological differences in their phonemic inventories and prosodic systems.[30–34] Some phonemes are unique to either Basque or Spanish. For example, Spanish has a voiceless velar fricative, while in Basque, this sound is in allophonic alternation with a voiced palatal approximant, and the distribution of dominant variants is highly variable, compromising the contrastive status of this phone. Basque has a set of phonemes that are not used in Spanish, including voiceless alveolar affricates, voiceless palatal fricatives, and so on. Opposition between /b/ and/v/ in Basque is contrastive, while in Spanish, /b/ versus /v/ opposition is allophonic, not contrastive (the same is true of /d/ and /g/ phones, which do not contrast with their corresponding—by place of articulation—fricatives). Basque and Spanish also differ in terms of unmarked locations of lexical stress and in their inventories and implementation of phonological tones. Catalan and Spanish have different numbers of vowels (eight in Catalan and five in Spanish), phonotactic constraints (Catalan allows more consonantal clusters than Spanish), and intonational patterns.[35–40] Bilinguals, who are constantly exposed to different languages, have to process a wider range of speech cues than monolinguals. Their rich linguistic immersion can lead to greater SL efficiency and may also prompt further adaptations of SL mechanisms for speech processing on the ontogenetic timescale.

There are also important differences in regional variation across languages. Catalan has only two main regional varieties: Eastern Catalan—spoken in Barcelona and the Province of Tarragona, which serves as a standard variety; and Western Catalan—spoken in the Autonomous Region of Valencia (to the east of the city of Tarragona), Lleida province, and the Balearic Islands (although the Mallorcan variety is often singled out as a separate regional dialect, intermediate between the other two). Spanish exhibits a broad dialectal continuum, from the standard Castilian Spanish variety and regional accents on the Iberian Peninsula to multiple Latin America varieties. Importantly, regional varieties of Spanish (and Catalan) are mutually understandable

Ann. N.Y. Acad. Sci. xxxx (2022) 1–19 © 2022 The Authors. *Annals of the New York Academy of Sciences* published by Wiley Periodicals LLC on behalf of New York Academy of Sciences

and share the same phonology (phonemic inventory and stress location); differences between regional accents of Spanish pertain solely to phonetics rather than phonology. In other words, the acoustic differences between regional Spanish accents relate to phonetic realization and are not systemic. In contrast, despite its relatively small geographical range, Basque features more regional variation than Spanish and differences between regional varieties of Basque are systemic. For example, Bizkaian varieties make lexical contrasts between accented and unaccented words (similar to those in Swedish and Japanese). Lapurdian accents, influenced by French prosody, use phrase-level accentuation and no word-level prominence or contrastive and lexical stress. Zuberroa, another Basque dialect spoken in France, has seven vowels instead of five, makes a phonemic contrast between nasal and oral vowels (due to the influence of French) and, like Lapurdi, only exhibits phrase-level prominence (no contrastive word-level stress). All Basque dialects also differ in terms of consonant inventories. Hence, Basque speakers are exposed to more dialectal variation and need to cope with systemic varietal differences, while Spanish speakers only have to cope with regional differences in pronunciation that are mostly limited to differences between regional accents in phonetic implementations. Hence, Basque bilingual populations may benefit from regional variation and show processing advantages due to training on a wider range of speech cues. This could lead to further adaptations of SL on the ontogenetic timescale, as individuals learn to cope with a bilingual environment featuring multiple systemic differences between regional varieties.

In our study, we used the recognition test (specifically, postfamiliarization recognition accuracy measured as the number of correct responses) from the classic artificial language learning paradigm to test for SL efficiency. In the auditory modality, we expected all groups would demonstrate higher accuracy on linguistic than nonlinguistic materials and that both bilingual groups would show an additional advantage on linguistic material, demonstrating the influence of bilingual speech input on the ontogenetic timescale. In the visual modality, we expected better recognition accuracy on nonlinguistic material across groups (replicating our earlier results). We did not have clear predictions as to whether simultaneous exposure to multiple languages would affect recognition accuracy in the visual modality, but intended to explore this possibility using between-group comparisons. It was possible that bilinguals' experience with different languages would also modulate SL in the visual modality, reflecting the effects of written language experience on individual development. If so, we expected to observe differences between the monolingual Spanish and bilingual Catalan and Basque participants. This would indicate that ontogenetic factors can lead to adaptations of SL cognitive mechanisms that remained relatively stable on the phylogenetic timescale.

Materials and methods

We reused the linguistic and nonlinguistic material from Ordin *et al.*,[29] but did not include the semilinguistic material. The linguistic material comprised a set of syllables. In the auditory modality, 32 consonant-vowel syllables were synthesized from Spanish phonemes; 24 of these syllables were used to create eight triplets: /ko-fa-me/, /fo-na-ku/, /mo-si-ke/, /ka-so-ni/, /sa-mu-pe/, /no-su-pi/, /po-fu-mi/, and /fe-nu-pa/. Within triplets, each syllable predicted the following syllable with 1.0 probability, that is, TPs between adjacent syllables within triplets were all set to 100%. In natural languages, content words are usually combined with pre/postpositions, articles, and clitics into phonological words,[41] and such functional elements can be modeled in artificial languages by "filler" syllables between content words.[42] We used the eight remaining synthesized syllables—/ma/, /fi/, /pu/, /se/, /ne/, /ki/, /li/, and /lu/—as functional elements. We used only open (CV, i.e., consonant-vowel) syllables because they are cross-linguistically unmarked, that is, if a language has syllables with codas (i.e., consonants or consonantal clusters after the vowel), it also necessarily has open syllables, yet the reverse is not necessarily true. Hence, open syllables with a single consonant before the vowel are universal and do not provide processing advantages for native speakers of any language. In all the native languages of our participants—Basque, Catalan, and Spanish—open syllables (CV) predominate, although Catalan allows for more complex clusters.[38]

Using these materials, we implemented the following hierarchical linguistic structure. We first concatenated the triplets and fillers using the grid

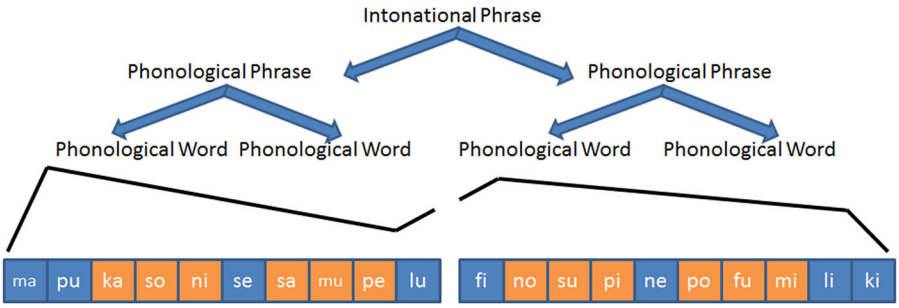

Figure 1. **A schematic representation of a familiarization (exposure) stream. Different shades of gray (in print) or colors (online) represent filler syllables (blue) and triplets (orange). Imposed F0 contour is represented by the straight lines over the syllabic grid. Prosodic hierarchy is displayed on the top. The figure shows one intonational phrase from a familiarization stream.**

(Fig. 1) to create artificial *phonological words*. TPs between these fillers and triplet-boundary syllables were approximately 12.5%. Then, we concatenated pairs of phonological words to create *phonological phrases* (PPs). We then concatenated pairs of PPs to create *intonational phrases* (IPs; or sentences). The structure of the IP is displayed in Figure 1. Finally, we concatenated IPs to produce the exposure stream. Overall, each triplet was presented 80 times in the exposure stream, and we avoided repeating the same triplet within one IP prosodic frame. Each filler syllable was embedded an equal number of times during an exposure stream. The bilingual Basque–Spanish and Catalan–Spanish assistants checked that no real words from participants' native languages emerged in the exposure stream. The exposure stream was synthesized in MBROLA,[43] such that each syllable lasted 240 ms (140 ms—vowels), with 50 ms pauses between PPs within IPs and 150 ms between IPs. An intonational contour, with a declination trend from 210 to 100 Hz, was imposed on each IP; the last syllable of each IP was marked by a sharp falling tone that dropped from 100 to 80 Hz. Rising tones were then added to mark the beginning and end of each PP. These acoustic parameters simulate the prosody of natural languages.[44,45] Coarticulatory information that might be relevant for detecting word boundaries in natural speech was not included because we wanted to ensure that participants could only use TPs to segment triplets. Note that, because of fillers, triplets did not start immediately after or finish with a pause. Boundary tones were implemented on filler syllables and thus did not cue triplet edges. This was done to prevent listeners from using acoustic cues to detect triplet edges; they were forced to rely only on statistical computations (or frequency of co-occurrence of syllables) for segmentation (see the description of the task below).

For nonlinguistic material, we used 32 transient sounds (water drops, footsteps, squeaks, animal noises, etc.) from https://freesound.org. These sounds were equalized (compressed or extended) to last 300 ms and normalized in intensity to 80 dB (to make them perceptually similar in terms of loudness). Of these sounds, 24 were used to make eight triplets and eight were reserved for fillers. The sounds were concatenated in the same statistical structures, using the same grid, as the syllable exposure stream. The duration of pauses between larger constituents was set to 200 ms, while pauses between shorter constituents were set to 100 milliseconds. Instead of the F0 modifications (implemented in the linguistic material), we decreased linear intensity to provide a boundary cue at the end of the larger prosodic constituents, and increased linear intensity to cue the beginning of constituents. Amplitude ramping was applied over the initial and two final syllables using the "filler" sounds at the edges of constituents, that is, those sounds that are not included in the recurrent triplets. These prosodic manipulations are not typical of natural speech, but provided structural complexity similar to that implemented in the linguistic material.

For the recognition test, we synthesized isolated triplets that were embedded in the exposure streams (linguistic: F0 set to 120 Hz, monotone, 240 ms syllables, 140 ms vowels; nonlinguistic: 300 ms sounds, normalized by intensity level). The syllables/sounds used in the recurrent triplets were redeployed to create foils (sequences of three syllables/sounds in which the TPs between elements within triplets

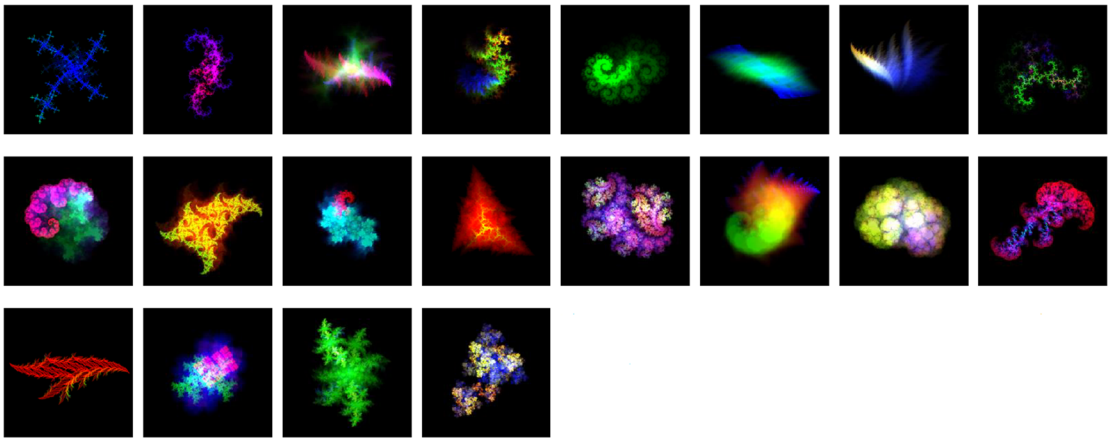

Figure 2. **Fractals used for the nonlinguistic material. Fractals in the upper row were used as fillers; fractals in the lower two rows were used to compose triplets.**

were violated). Two elements that appeared consecutively in the foils were never used consecutively in the triplets. Filler elements from the exposure streams were not used in the recognition test tokens. In one set of eight foils, the order of elements was preserved (if a particular syllable/sound had been used in the triplet-initial position in the exposure stream, it was also in the foil-initial position). In the other set of eight foils, the order of elements inside triplets was violated (e.g., a triplet-initial syllable/sound would only appear in the foil-medial or foil-final position). Sample test tokens and familiarization stream in linguistic and nonlinguistic domains can be found in File S1 (online only).

In the visual modality, we used syllables for the linguistic material and fractals (Fig. 2) for the nonlinguistic material. We used 12 syllables to compose four triplets (TE-GU-BA, TA-BO-FA, KA-BE-TO, and GA-FO-BU), and eight syllables for the as fillers (TU, GO, GE, KO, KU, FU, FE, and KE). The fillers and triplets were then arranged according to the same grid as used for the auditory modality. Importantly, we used different syllables in the auditory and visual modalities. The elements were then presented in the middle of the screen, one by one, for 500 ms each. Boundary cues, displayed for 500 ms each, were punctuation marks (commas and periods) in the linguistic material and empty squares (white–prominent against the dark gray background or lighter gray–one tone lighter than the background) in nonlinguistic materials. Each triplet was presented 50 times in an exposure stream. For the recognition test, we used four triplets and eight foils each from the linguistic and nonlinguistic domains. Ordinal positions within the foil were preserved in half but violated in the other half of the foils.

To ascertain to what extent peculiarities of the linguistic and nonlinguistic material might drive effects, and whether stimuli in different domains and modalities were equally discriminable, we ran a norming verification study with 24 participants (age range: 18–35, $M = 24$) in the Basque country who did not take part in the main experiment. Half reported they were Basque–Spanish bilinguals, the other half reported they were newcomers to the Basque country who had no previous exposure to the Basque language.

The norming experiment comprised two sessions. In session 1, we used the same syllables, fractals, and sounds that composed the recurrent triplets in the main experiment: auditory linguistic syllables (240 ms each); visual syllables (500 ms); auditory nonlinguistic sounds (300 ms); and visual fractals (500 milliseconds). On each trial, participants saw three stimuli from the same category; the second or the third stimulus was identical to the first. There were buttons with marks "2" or "3." Participants had to decide whether the second or the third stimulus was identical to (50% of trials) or different from (50% of trials) the first stimulus by pressing the corresponding button with a mouse. The task message indicating whether they should choose identical or different stimulus from the first one in the sequence appeared only after all three

stimuli were presented (because we wanted people to pay equal attention to all three stimuli). Each target stimulus appeared once in the similarity and once in discrimination task, for a total of 48 auditory trials and 24 visual trials per experiment. The order of the four experiments was counterbalanced so that 12 participants first did an experiment with linguistic stimuli (six auditory first; six visual first) and 12 participants started with nonlinguistic stimuli (six auditory first; six visual first). The order of stimuli within experiments was randomized for each participant. Session 2 included four similar experiments, except that the stimuli were not separate syllables or sounds, but whole triplets. They completed 16 trials in the audio modality and eight trials in the visual modality. Audio triplets were separated by a 1000 ms pause; visual triplets by a blank screen. We used the same counterbalancing procedure for experiments and per-participant randomization of trials in Session 2.

All participants performed at ceiling level, responding correctly on all trials in all norming experiments, suggesting that auditory syllables, sounds, visual syllables, and fractals were all equally and easily discriminable from one another, both individually and when combined into triplets.

Procedure

We used a 2∗2 within-subject experimental design with two sessions. In one session, participants performed two experiments in the visual modality (linguistic and nonlinguistic material); in the other, two experiments in the auditory modality (linguistic and nonlinguistic material). The order of sessions and the order of experiments within sessions were counterbalanced across participants.

Each experiment consisted of a familiarization and a recognition section. During familiarization, people were exposed to a familiarization stream. For nonlinguistic material, we asked participants to watch alternating fractals displayed in the middle of the screen or listen to a long stream of sounds, and to detect and memorize recurrent sequences of fractals or sounds. For linguistic material, participants were informed that they will learn an alien language by listening to speech or see a text presented syllable-by-syllable in the middle of the screen; their task was to detect and memorize the words of the alien language. A yes/no recognition test was administered after each familiarization section. The participants saw either triplets from the familiarization stream or foils and had to say whether they had seen these during familiarization (for nonlinguistic material) or whether these were words from an alien language (for linguistic material). On each trial, we asked participants how sure they were that their response is correct (on a 4-point scale). Confidence ratings were collected for a different study, and not analyzed here. Each test token was used twice (yielding 48 trials in the auditory modality per test and 24 trials in the visual modality per test). The order of trials was randomized individually for each participant.

After the last experiment, participants performed a subset of the Spanish KBIT IQ test[46] targeting only logical IQ, which has been shown to be valid and to capture individual differences reliably across research and clinical contexts.[47,48] Then, participants did a rapid picture-naming test (bilingual participants in both of their native languages and monolinguals in Spanish). The test was based on the multilingual lexical test introduced by Gollan *et al.*[49] It comprised 65 images representing common entities from different categories (animals, body parts, and everyday objects), which were noncognates in both language pairs (e.g., *mesa* (Spanish); *taula* (Catalan); *mahaia* (Basque); gloss: *table*). Using instructions in the corresponding language, we asked participants to name all the objects first in one of the languages, then in the other (order counterbalanced across participants). Finally, bilinguals filled in a language background questionnaire, detailing the age of acquisition for both languages; percentage of time each language was used in various social contexts, for example, communication with a partner/parents/children/friends/colleagues/educators, and so on; their level of proficiency in other foreign languages; and their preferred language for leisure, that is, reading and movies.

Participants

We recruited 39 native Spanish monolingual speakers residing in a monolingual area (the city of Teruel), age range: 18–35, $M = 24$; 43 Catalan–Spanish bilinguals (a Western Catalan variety typical of Valencia), age range: 18–35, $M = 25$; and 46 Basque–Spanish bilinguals (standard Basque—Batua—from Gipuzcoa), age range: 18–35, $M = 25$. The sample of Basque bilinguals was different from that reported in Ordin *et al.*[29] All participants were

matched in socioeconomic status and did not have speech/language/hearing/cognitive disorders. We did not recruit regular users or proficient speakers of other languages (e.g., students of modern languages and translation; anyone who reported being a frequent and fluent speaker of other languages was excluded from the sample). All bilingual speakers had been exposed to a bilingual environment from birth and had acquired both languages simultaneously. They also reported being more frequent users of their minority (Basque or Catalan) language in informal social contexts and using both languages equally in formal social settings.

We used the KBIT test to describe the samples in terms of their IQ score distribution. The IQ scores within each group were distributed normally (Shapiro–Wilk test: $P > 0.02$ for each group) and no differences in variance of IQ score distributions were observed (Levene's test: $P = 0.158$). An analysis of variance (ANOVA) did not reveal significant differences between the group means, $F(2,124) = 0.209$, $P = 0.812$.

The rapid naming test, a proxy for lexical access efficiency, was used to assess the relative proficiency of bilinguals in each of their languages. For each image correctly named in a corresponding language, participants received one point (the maximum score is 65 points per language). Smaller individual differences reflect more balanced bilingual proficiency. In Spanish, all bilinguals (and monolinguals) achieved maximum scores. However, the Basque participants scored higher in Basque (median $M = 64$) than Catalan participants scored in Catalan ($M = 61$). A Mann–Whitney U-test showed a significant difference in lexical access in minority languages between bilingual groups, $W = 1636.5$, $P < 0.001$, $\Delta M = 3$ (Fig. 3).

This difference was due to the fact that Catalan bilinguals often used a Spanish noun to name an object in the rapid picture-naming task, and such responses were considered incorrect. Basque bilinguals did not use Spanish words to name objects in the same test administered in Basque. This difference can probably be attributed to greater overlap in the Catalan–Spanish than the Basque–Spanish bilingual lexicons. Both Catalan and Spanish are Romance languages and share most vocabulary stems. Substituting a Spanish word for a Catalan word or vice versa can often go unnoticed in everyday conversations. By contrast, Basque and Spanish

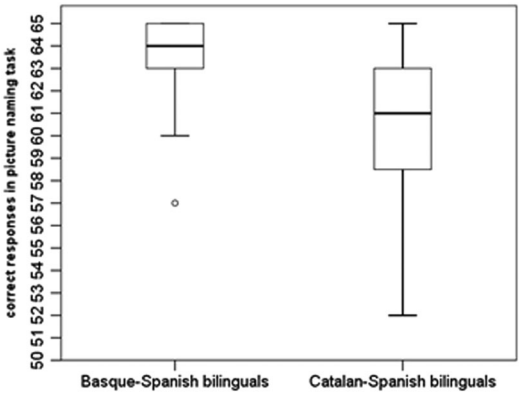

Figure 3. **Number of correct responses in Catalan (for Catalan–Spanish bilinguals) and in Basque (for Basque–Spanish bilinguals) in the picture-naming task. The box shows the middle 50% range; the horizontal line inside the box shows the median; the bars represent the first and fourth quartiles. The range of possible scores is between 0 and 65 (minimum and maximum scores).**

are genealogically diverse languages, and differences in core vocabulary may prevent bilinguals from substituting Basque for Spanish words, thus leading to the consolidation of two discrete vocabulary items. Moreover, Basque and Spanish are typologically different, hence using a Spanish word in a Basque sentence often requires adaptations to morphology. For example, using a Basque word in a Spanish sentence may require deciding how to assign noun gender in Spanish (Basque does not have gender markers). In short, Basque–Spanish bilinguals face various cross-language challenges, which may lead to greater separation between their mental lexicons. Hence, the slight difference in test scores and the large difference in score variability do not necessarily indicate that Catalan–Spanish bilinguals had less balanced proficiency in their languages, but may instead reflect the differing degrees of separation in their mental lexicons.

Analysis

We were interested in the efficiency of SL between domains (linguistic versus nonlinguistic) in two perceptual modalities (visual and auditory). Efficiency can be related to multiple parameters: (1) speed of learning; how fast regularities are detected and discrete patterns are extracted from a continuous sensory input; (2) automaticity; how much cognitive resources are consumed by SL and to what extent SL is compromised by parallel tasks

that divert attentional or memory resources; (3) stability of SL; the steepness or speed of the learning decay; and (4) transferability; whether extracted regularities can be used to better process new inputs, unrelated to those from which these regularities were extracted. Within the framework of this project, we measured SL efficiency in terms of accuracy (number of correct responses) in the recognition test. Correctness, however, can also be determined independently by how well false patterns are rejected, and how well correct patterns are accepted (hence we explored rejection and acceptance accuracy separately).

Our analysis uses signal detection theory (SDT) and the false discovery rate (FDR). The latter is common in evaluating the efficiency of pattern recognition algorithms. This means we consider a human participant as an information processing unit who has to sort the test tokens presented into relevant (i.e., triplets) and irrelevant (i.e., foils) categories. As in all binary classification tasks, accepted triplets are considered *true positives*; accepted foils, *false positives*; rejected triplets, *false negatives*; and rejected foils, *true negatives*. The efficiency of decision-making strategies adopted by a participant was evaluated using *sensitivity* (the ratio of true positives to all other relevant tokens, i.e., the proportion of the relevant tokens that were endorsed) and *specificity* (the ratio of true negatives to all incorrect responses) measures (Fig. 4). In other words, sensitivity is the probability of detecting relevant tokens among all available tokens; and specificity is the probability of identifying irrelevant token as relevant tokens.

Within the SDT analytical framework, true positives are defined as hits, while false positives are defined as false alarms. Unselected relevant items (those outside the large circle in the middle of Fig. 4) are defined as misses, and unselected irrelevant items are defined as correct rejections. Unlike SDT, FDR can measure the efficiency of rejection and efficiency of endorsement separately. Since we believe that from an evolutionary perspective SL was honed by the need to detect breaks in statistical structure, it is highly possible that higher efficiency due to evolutionary selection pressures will be reflected in better specificity. Besides, we have two different types of foils, and the FDR approach allows us to focus on these two types of foils separately so as to evaluate the respective contributions of positional memory and statistical regularity to the efficiency of detecting foils. In fact, our choice of the analytic approach reflects the importance of distinguishing between the efficiency of triplet endorsement and foil rejection. Higher values of sensitivity and specificity stand for higher SL efficiency in terms of acceptance and rejection accuracy, respectively.

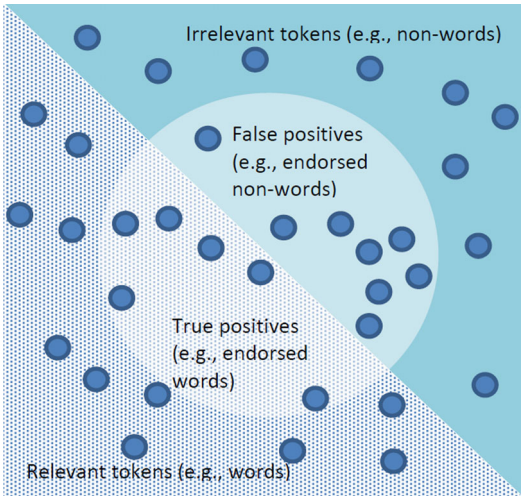

Figure 4. **The large square shows the population of tokens that are divided into binary classes by the participants (e.g., into words and nonwords). As participants need to respond whether an item is a word or not, the nonwords are irrelevant items, and the words are relevant items. The large circle in the middle embraces the tokens that are endorsed as words by a participant. Endorsed irrelevant items are false positives, and endorsed relevant items are true positives.**

The efficiency of the classification algorithm, however, depends on a combination of sensitivity and specificity, taking any possible bias to either accept or reject presented tokens into consideration (i.e., individual overall tendency to respond "yes" or "no"). To measure efficiency in terms of combined rejection/acceptance accuracy, we first calculated a *precision* value for each individual as a ratio of true positives to all endorsed tokens (the probability that an endorsed token was indeed a triplet). Precision gives us the probability that a given token was a triplet from the exposure input, given that it was selected by an individual. This allows us to calculate D-prime (d′), a measure of individual sensitivity that is unaffected by any individual bias to accept or reject tokens. D-prime provides a different measure of efficiency (in terms of accuracy) that can be considered as a verification of the interpretation based on the FDR precision measure.

Due to multiple differences in the subsets of SL mechanisms that operate in the visual and auditory modalities,[22,50,51] direct comparison of SL efficiency between modalities is not meaningful. Instead, we compare FDR measures across linguistic and nonlinguistic material (within-subject) and between groups (Basque–Spanish, Catalan–Spanish, and monolingual Spanish participants).

Results

Comparing the groups

IQ scores were not different between linguistic populations, but the scores on rapid naming tests in Catalan by Catalan–Spanish bilinguals were significantly lower than those in Basque by Basque–Spanish bilinguals. In the description of our linguistic populations in the Methods section above, we suggested that lower scores in the picture-naming task by Catalan bilinguals did not reflect lower proficiency in their minority (i.e., local) language but rather the relative ease of exchanging Catalan and Spanish words within a single sentence. However, additional tests were run to verify whether this difference between bilinguals could influence accuracy in the recognition tests. Although the scores on the rapid naming test showed high proficiency both in Basque and Catalan for the corresponding bilingual samples, the variance of the scores was larger for the Catalan (range: 52–65) than for the Basque (range: 60–65) speakers. We performed a median split on the Catalan speakers, based on their rapid naming test scores (median score = 61; 18 participants had lower scores in the range 52–60; 21 participants had higher scores in the range 62–65; and 3 participants had median scores). Next, we ran *t*-tests on each measure of SL efficiency (*D-prime, precision, recall, bias,* and *overall specificity,* separately in the auditory and visual modalities). We did not observe any significant differences between Catalan participants with lexical test scores below or above the median. Levene's test showed that the variance in the values of the dependent variable was equal in the subgroups formed based on the median split. Spearman correlations between each dependent variable and scores on the lexical test, as well as all correlations, proved insignificant and negligible (all test results can be found in the File S2, online only). These results strengthen the argument that the variability of lexical test scores had no effect on SL efficiency. This variability probably resulted, as we suggested earlier, from the fact that Catalan and Spanish are genealogically close so that a word from one language can be used in place of a word in the other language; Catalan speakers tended to name objects in Spanish during the test, as they would do it in their utterances in Catalan, but technically these responses had to be considered wrong answers, decreasing test scores.

Auditory modality

D-prime (auditory). In our earlier experiment, we found that d′ in the auditory modality was higher on linguistic than on nonlinguistic material in the group of Basque–Spanish bilinguals. In this study, we wanted to compare the difference in d′ between the three populations. For this purpose, we calculated d′ by taking the difference between d′ on linguistic and nonlinguistic material for each individual, then performed the analysis of variance (ANOVA, assumptions of normality, and equal variance are controlled for by the Shapiro–Wilks and Levene's tests) on individual deltas with *group* as a factor. The analysis showed that differences in d′ varied significantly between groups, $F(2,125) = 3.427$, $P = 0.036$.

We hypothesized that the effect of a bilingual environment would lead to adaptations of SL mechanisms in the ontogenetic timecourse of individual development and increase the difference in performance on linguistic versus nonlinguistic material in bilinguals relative to monolinguals. Pairwise comparisons (all P values corrected by the Holm–Bonferroni method here and below, unless otherwise specified) showed that Δd′ was larger in the group of Basque bilinguals than in the group of monolinguals ($P = 0.036$). However, we did not observe significant differences in Δd′ between Catalan bilinguals and Spanish monolinguals ($P = 0.861$). Interestingly, the difference in Δd′ was significant between Basque and Catalan bilinguals ($P = 0.02$). The pattern suggests that the Basque–Spanish bilingual environment increased the difference in SL performance on linguistic versus nonlinguistic material compared to both the monolingual and Catalan–Spanish environments.

Precision (auditory). Precision values—positive predictive values—reflect the probability that an endorsed token is an actual triplet from the familiarization input (Fig. 5). ANOVA (with *material type* as within-subject factor and *group* as

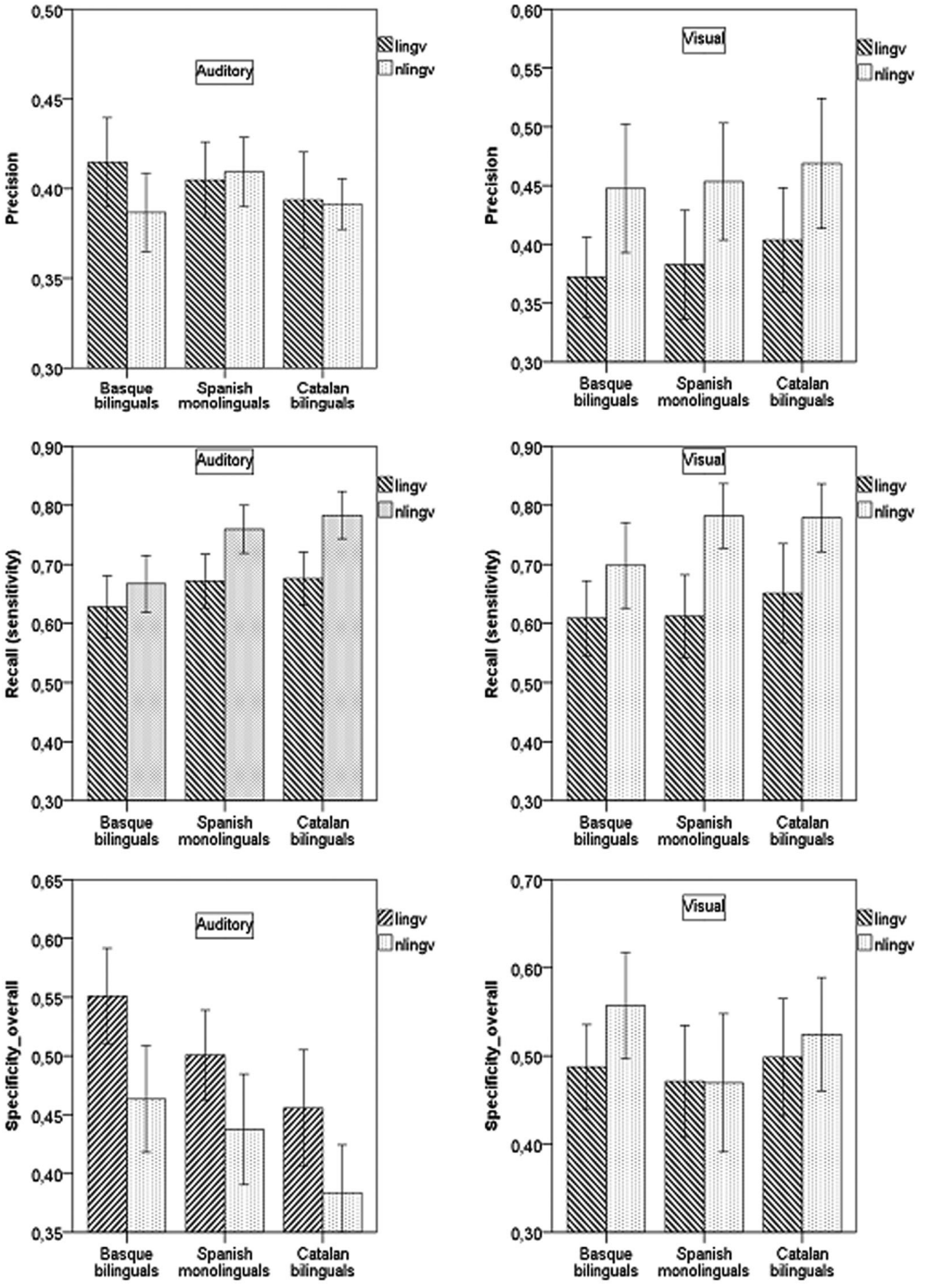

Figure 5. **Precision, sensitivity (acceptance accuracy), and specificity (rejection accuracy) on linguistic and nonlinguistic stimuli in the visual and auditory modalities. Specificity graphs are given for overall specificity, across both types of foils. Error bars show 95% CI.**

between-subject factor) did not reveal statistical differences either between groups, $F(2,125) = 0.715$, $P = 0.491$, or between material types, $F(1,125) = 1.174$, $P = 0.281$. There was also no significant interaction between *group* and *material type*, $F(2,125) = 1.595$, $P = 0.207$.

Sensitivity (recall) (auditory). Sensitivity (recall), the *true positive rate*, reflects the probability that presented tokens will be endorsed if they are indeed recurrent triplets from the familiarization input (Fig. 5). We subjected the recall values to ANOVA with *group* as a between-subject factor and *material type* as a within-subject factor. The analysis revealed a significant effect of *material type*, $F(1,125) = 29.157$, $P < 0.0005$, with better recall on nonlinguistic than linguistic material in all samples, and *group*; $F(2,125) = 5.529$, $P = 0.005$. There was no significant interaction between *material type* and *group*, $F(2,125) = 2.05$, $P = 0.133$. Pairwise comparisons showed that recall was lower in the group of Basque–Spanish than in Catalan–Spanish bilinguals ($P = 0.04$) or Spanish monolinguals ($P = 0.007$), with no difference between Catalan bilinguals and monolinguals ($P = 1.0$). Thus, nonlinguistic triplets are recognized and retrieved better than linguistic ones, and overall Basque bilinguals miss more triplets than Catalans or monolinguals. At the same time, the percentage of correctly retrieved tokens among all retrieved tokens did not differ between groups. These results for precision and sensitivity can be explained by a stronger tendency to accept tokens—irrespective of their correctness—by Catalans and monolinguals. Hence, we next compared the bias (criterion C) between groups.

Bias (C) (auditory). All individuals had a strong tendency to accept tokens, which is not surprising given that the recognition test included twice as many foils as triplets. An ANOVA with *group* and *material type* showed that participants were more likely to accept nonlinguistic test tokens than linguistic ones, $F(1,125) = 43.371$, $P < 0.0005$. The effect of *group* was also significant, $F(2,125) = 6.289$, $P = 0.002$, with no significant interaction between *group* and *material type*, $F(2,125) = 0.926$, $P = 0.399$. Pairwise comparisons showed that overall Catalan bilinguals had a higher tendency to accept tokens than Basque bilinguals, $P = 0.002$. Other pairwise differences between groups were not significant. Recall was significantly different between Basque bilinguals and Spanish monolinguals, but they did not differ on acceptance bias, suggesting that the combination of recall and precision patterns could not be fully accounted for by response bias.

Specificity (auditory). Specificity—the true negative rate—reflects the probability that test tokens are not endorsed if they are foils (participants are good at rejecting tokens that were not embedded into the familiarization input). The ANOVA with *group* and *material type* as factors showed a significant effect of *material type*, $F(1,125) = 37.333$, $P < 0.0005$ and *group*, $F(2,125) = 5.492$, $P = 0.005$, with no interaction between the *groups*, $F(2,125) = 0.347$, $P = 0.707$. Pairwise comparisons showed that the only significant pairwise difference was between Catalan and Basque bilinguals, who had a higher true negative rate, $P = 0.004$ (Fig. 5).

We then separately compared specificity for different foil types. First, we calculated specificity for foils, in which the positional order of elements (syllables or sounds) was preserved, and subjected these measures to an ANOVA with *material type* and *group* as factors. We observed a significant effect of *material type*, $F(1,125) = 12.57$, $P = 0.001$, with higher specificity on linguistic material. We found no significant effect of *group*, $F(2,125) = 2.344$, $P = 0.1$, and no significant interaction between the factors, $F(2,125) = 0.877$, $P = 0.419$. For the foils, in which the ordinal position of elements was violated, we observed a significant effect of *material type*, $F(1,125) = 34.654$, $P < 0.0005$ (specificity was higher on linguistic material) and *group*, $F(2,125) = 7.97$, $P = 0.001$, but no significant interaction between these factors, $F(2,125) = 0.66$, $P = 0.519$. Pairwise comparisons for the *group* factor reveal that specificity on such foils was higher in Basque than Catalan bilinguals, $P < 0.0005$, and in Catalan bilinguals than Spanish monolinguals, $P = 0.036$, with no significant difference between two groups of bilinguals (Catalan and Basque). This pattern shows that rejection accuracy (the true negative rate) based on detecting violations of statistical regularities did not differ between groups. However, Basque bilinguals were better at rejecting foils that violated both statistical regularities and positional order, suggesting that the contribution of memory to rejection of statistical regularities was lower in the Catalan bilinguals than in the other two groups.

Visual modality

D-prime (visual). ANOVA with *material type* and *group* as factors showed that d′ was significantly higher on nonlinguistic than linguistic stimuli, $F(1,125) = 26.98$, $P < 0.0005$ but *group*, $F(2,125) = 1.325$, $P = 0.27$, and the interaction between factors, $F(2,125) = 0.01$, $P = 0.99$, were not significant. We then calculated Δd′ by taking the difference between d′ on linguistic and nonlinguistic material for each participant and performed an ANOVA on the differences in d′ as a factor of *group*. The analysis did not reveal a significant effect of *group*, $F(2,125) = 0.01$, $P = 0.99$. Overall, this pattern of results confirms that in the visual modality, SL was more efficient on nonlinguistic material, and this effect was not modulated by individual linguistic experience.

Precision (visual). We ran an analysis on positive predictive values with *group* and *material type* as factors. The analysis showed a significant effect of *material*, $F(1,123) = 15.05$, $P < 0.0005$, with no significant effect of *group*, $F(2,123) = 0.512$, $P = 0.601$, and no significant interaction between *group* and *material type*, $F(2,123) = 0.043$, $P = 0.957$. We observed higher precision on nonlinguistic stimuli than linguistic stimuli in all groups (Fig. 5). It appears that all people are more likely to endorse only relevant tokens on nonlinguistic material.

Sensitivity (recall) (visual). Recall is higher on nonlinguistic than linguistic stimuli across all groups (Fig. 5). The analysis showed a significant effect of *material type* on recall, $F(1,125) = 23.186$, $P < 0.0005$. The effect of *group*, $F(2,125) = 1.731$, $P = 0.181$, and interaction between *group* and *material type* were not significant $F(2,125) = 0.741$, $P = 0.479$.

Specificity (visual). We ran an analysis on the true negative rate scores with *group* and *material type* as factors. This analysis did not reveal any significant effect of *material type,* $F(1,125) = 2.063$, $P = 0.153$, *group* $F(2,125) = 1.151$, $P = 0.32$, or any interaction between *group* and *material type,* $F(2,125) = 0.904$, $P = 0.407$. This suggests that there was no difference in the efficiency with which participants with different native languages rejected linguistic and nonlinguistic foils (Fig. 5).

Next, we compared specificity for different foil types. In rejection efficiency for foils in which the positional order of elements was preserved, we did not observe a significant effect of *material type* $F(1,125) = 0.063$, $P = 0.802$ or a *material type* by *group* interaction $(2,125) = 0.115$, $P = 0.891$. The effect of *group* was significant, $F(2,125) = 3.246$, $P = 0.042$. Pairwise comparisons revealed no significant differences between bilingual groups, $P = 0.815$, but higher specificity in the group of Basque bilinguals compared to monolinguals, $P = 0.02$ (uncorrected), and in the group of Catalan bilinguals compared to monolinguals, $P = 0.039$ (uncorrected). However, after applying correction for multiple comparisons, the P values failed to exceed the *a priori* α threshold (0.05). For the foils, in which the ordinal position of elements was violated, we observed a significant effect of *material type*, $F(1,125) = 4.709$, $P = 0.032$, with high specificity on nonlinguistic material. The effect of *group*, $F(2,125) = 0.075$, $P = 0.928$, and the interaction between *material type* and *group*, $F(2,125) = 2.528$, $P = 0.084$ were not significant. This pattern of results allows us to make only one unambiguous conclusion: rejection efficiency is modulated by material type only when the positional order of elements is violated, with better rejection accuracy on nonlinguistic material.

Bias (C) (visual). All individuals exhibited an overall tendency to accept tokens, which was modulated by *material type*, $F(1,125) = 6.011$, $P = 0.016$, with no significant effect of *group* on the endorsement bias, $F(2,125) = 1.308$, $P = 0.274$, or interaction between *material type* and *group* $F(2,125) = 1.174$, $P = 0.313$. This endorsement bias was stronger for nonlinguistic than linguistic material. This might explain the higher proportion of endorsed relevant tokens (i.e., *recall*) in the nonlanguage compared to the language domain across all groups. However, it does not account for equal specificity across all groups and material types (in the visual modality) or *precision* (the proportion of relevant tokens among all endorsed tokens). These results cannot be completely explained by the stronger overall tendency to endorse nonlinguistic tokens.

Discussion

Our analyses aimed to compare SL performance on linguistic and nonlinguistic material in the visual and auditory perceptual modalities, with *material*

Table 1. Summary of significant results

Measure		Auditory results		Visual results	
Δd′	$d'_{ling} - d'_{nonLing}$: Is sensitivity within modality higher on linguistic or non-linguistic material?	*Basq>Cat*; *Basq>mono*		n/s	
C	Bias criterion: Are people more likely to endorse presented tokens (positive C values) or reject them (negative C values)?	Material	*NonLing>Ling*	Material	*NonLing>Ling*
		Group	*Basq<Cat*	Group	n/s
Precision	Proportion of relevant tokens among all endorsed tokens:	Material	n/s	Material	*NonLing>Ling*
	Do people retrieve well?	Group	n/s	Group	n/s
Sensitivity (recall)	Proportion of endorsed tokens among all relevant tokens:	Material	NonLing>Ling	Material	*NonLing>Ling*
	Do people retrieve well?	Group	*Basq<Cat*; *Basq<mono*	Group	n/s
Specificity	Proportion of foils that were not endorsed:	Material	*Ling>NonLing*	Material	n/s
	Do people reject well?	Group	*Basq>Cat*	Group	n/s
$Specificity_{Same\_Pos}$	Do people reject the foils that preserve the positional order of elements well?	Material	*Ling>NonLing*	Material	n/s
		Group	n/s	Group	*Basq>mono (uncorrected)* *Cat>mono (uncorrected)*
$Specificity_{Dif\_Pos}$	Do people reject the foils that violate the positional order of elements well?	Material	*Ling>NonLing*	Material	*NonLing>Ling*
		Group	*Basq>Cat* *mono>Cat*		

type (linguistic–*Ling* and nonlinguistic–*NonLing*) as a within-subject factor and *group* (Basque–Spanish bilinguals–*Basq*, Catalan–Spanish bilinguals–*Cat*, and Spanish monolinguals–*Mono*) as a between-subject factor. As the underlying cognitive processes of SL across modalities are different and differences in complexity of visual and auditory material are unavoidable, direct comparison between modalities was not carried out. Table 1 presents the main result patterns for each modality.

In our earlier study,[29] we found that sensitivity (measured as d′) in the population of Basque bilinguals was higher on linguistic material in the auditory modality and on nonlinguistic material in the visual modality. Here, we analyzed whether the difference in sensitivity on different types of material was further modulated by the native language of the participant. We found that d′ was higher on nonlinguistic material in the visual modality in both groups of bilinguals and in monolinguals, and this difference was not modulated by the native language(s) of the participants. In the auditory modality, Basque bilinguals were more sensitive to linguistic material than Catalan bilinguals and monolinguals. Building further on our hypothesis that long-term exposure to spoken speech on the timescale of natural evolution could have resulted in adaptive changes of SL mechanisms for processing linguistic information, we now suggest that linguistic experience at an ontogenetic timescale can modulate this enhancement. Long-term individual exposure to the Basque–Spanish bilingual environment hones SL in the auditory modality for speech processing. In the visual modality, we did not observe evidence that SL mechanisms, which conceivably evolved for processing fitness-related environmental information, have undergone adaptive changes at either the phylogenetic or ontogenetic

timescales, in order to afford better processing of linguistic information (this pattern of results might differ in deaf populations using sign language). This confirms our suggestion that written language, as a relatively recent cultural invention, has not yet influenced the underlying cognitive machinery of SL.

The advantage of the signal detection theoretic approach taken here is that it provides a sensitivity measure (i.e., d′) that is not biased by an individual's tendency to endorse or reject presented tokens (i.e., C). One of the project objectives was, however, to investigate recognition accuracy for old items (true endorsement rate) and novel items that were not implemented in the exposure input. Hence, we also used an FDR analytic approach adopted from machine learning and analyzed the efficiency of endorsements and rejections separately.

We observed that retrieval (endorsement of relevant tokens from the pool of presented tokens), or recall, was higher on nonlinguistic material in both modalities but in the audio modality, recall was lower in Basque bilinguals than in Catalan bilinguals and Spanish monolinguals. The recall measure is not bias-free and could potentially be attributed to a stronger tendency to accept nonlinguistic tokens in both modalities, irrespective of whether these tokens are relevant. However, across-group differences in the auditory modality do not correspond to those observed for the bias measure: Basque bilinguals had lower recall than Spanish monolinguals, but the difference in levels of bias went in a different direction: Basque natives had a lower—albeit insignificantly so—general tendency to endorse tokens, suggesting that the recall pattern is not completely accounted for by bias.

Specificity—or rejection efficiency—was not significantly different between material types and groups in the visual modality and was higher on linguistic than on nonlinguistic material in the auditory modality, with higher rejection efficiency by Basque than by Catalan bilinguals. The specificity pattern in the auditory modality could be attributed to differences in bias. However, in the visual modality, the bias and specificity measures did not go in the same direction because participants in all linguistic populations revealed a stronger tendency to endorse nonlinguistic tokens, which did not lead to significant differences in specificity. This uncoupling between bias and specificity suggests that rejection accuracy is partially independent of the general tendency to respond "yes" or "no" in the recognition test. The bias cannot fully account for the observed across-modality and group differences in the rates of true positive (recall) and negative (specificity).

The precision rate reflects a combined contribution of endorsement and rejection efficiency to the accuracy of token classification as relevant (extracted from the exposure input) or irrelevant (not embedded in the exposure input). In the visual modality, recall was higher on nonlinguistic material and specificity was not significantly different between material types and groups; hence, precision values were determined only by the recall measure and repeated the recall pattern. In the auditory modality, a higher recall on nonlinguistic material and higher specificity on linguistic material neutralized differences in the precision between linguistic and nonlinguistic domains. Hence, we consider precision to be less informative for this dataset, and here focus on discussing rejection and acceptance separately, as proposed in our justification for using the FDR approach.

We proposed that SL mechanisms were shaped for detecting deviations in incidentally detected environmental patterns (and later redeployed for processing linguistic information), which means that on environmental nonlinguistic stimuli, we should expect higher rejection accuracy (specificity) on nonlinguistic material. The observed pattern, however, is not in line with this prediction. In the visual modality, the significant advantage for nonlinguistic material was observed in recall (acceptance accuracy), not specificity. In the auditory modality, we observed better recall on nonlinguistic material as well, and better specificity on nonlinguistic material. This indicates that detection of irrelevant auditory tokens that violated statistical regularities was better on speech-like material, and detection of statistically congruent tokens, both visual and auditory, is better on nonspeech-like material. We still observe overall higher SL efficiency on nonlinguistic material in the visual modality, but not because nonlinguistic foils are rejected with higher accuracy, but rather because nonlinguistic statistically congruent triplets are endorsed better. This suggests that people detect the presence of structure, not absence of structure, as we proposed. This is a worrisome deviation from

our theoretical premises, constructed based on the evidence from earlier studies[28,52]—although it is not central to this particular study—and it calls for further explanation. Below, we propose another possible explanation, which should be verified empirically in future studies.

Learning, including SL, can be accounted for by the free energy principle,[53] which states that modeling the world for efficient processing by cognitive systems relies on minimizing entropy (i.e., free energy variation) in the sensory input. Maximum reduction of entropy for SL mechanisms occurs when TPs are minimized (at transitions where the following element is least predictable, entropy becomes maximal; once this uncertainty is resolved, there is maximum reduction of entropy). This principle was more prominently reflected in the linguistic material in the auditory modality, resulting in better detection of statistical violations and thus higher rejection accuracy. In the visual modality, the same free energy principle explains the tendency of organisms to reduce the number of cognitive states they have to maintain by detecting *sensory patterns* (each pattern or set of patterns is associated with a particular cognitive state or phenotypic response), and to revisit these states multiple times,[54] solidifying an associated phenotypic response. Hence, in the visual modality, which is honed for detecting recurrent (spatial) patterns rather than recurrent sequences, we observed higher recall on nonlinguistic material. Processing nonlinguistic material in the EEA shaped phenotypic responses and cognitive states over a longer evolutionary timespan.

We also observed that the influence of long-term exposure to spoken speech at the evolutionary timescale was further modulated by the long-term effects of native language at the timescale of individual development. Note that the differences are between Basque–Spanish bilinguals and Catalan–Spanish bilinguals and between Basque–Spanish bilinguals and Spanish monolinguals. We did not observe differences in SL efficiency between Catalan–Spanish bilinguals and Spanish monolinguals. This result required further investigation given that we expected both bilingual groups to exhibit similar behavioral patterns and to differ from monolinguals (see Introduction). Our initial hypothesis was based on the phonological differences between Basque and Spanish and between Catalan and Spanish, yet the Catalan participants were recruited in Valencia, a province where the western dialect of Catalan is more widespread than the standard (or eastern) Catalan variety. Importantly, western Catalan is more similar to Spanish in terms of vowel frequency, vowel distribution, and durational contrasts.[55,56] Standard Catalan has vowel reduction, which reduces the range of vowels from seven phonemes in stressed positions to three in unstressed positions, and reduces the durations of unstressed vowels, thus leading to higher durational ratios of stressed to unstressed vowels and higher variability in the temporal distribution of vowels and interstress intervals. Western Catalan has five vowels in unstressed positions (like Spanish) and does not enhance durational contrasts between stressed and unstressed vowels,[55] in line with other Romance languages, including Castilian Spanish.[56] Western Catalan and Spanish both exhibit a tendency to *trochaic* foot grouping, while Eastern Catalan exhibits *iambic* metrical patterns.[39] Finally, Eastern Catalan sometimes uses (SV)O intonational phrasing pattern, along with a more common (S)VO phrasing, while Spanish and Western Catalan exclusively group components in (S)VO phrasing, limiting the range of possible intonational phrasing patterns.[35,57] These phonological differences are sufficiently prominent to enable prelinguistic babies from monolingual Catalan families (i.e., in which all family members use only Catalan at home and with their babies, thus minimizing Spanish speech input) to discriminate between Eastern and Western Catalan utterances, based on prosodic and segmental differences between the varieties.[58] Prosodic cues as well as the distributional and temporal properties of the vowels that contribute to particular rhythmic properties are employed for detecting the boundaries between linguistic constituents.[59–61] Thus, the cues for segmenting continuous speech into PPs and words are different in Western and Eastern Catalan, and in Western Catalan, these cues are similar to those in Spanish. Basque bilinguals use different segmentation cues for processing utterances in Basque and Spanish. Thus, exposure to a Basque–Spanish environment would do more to enhance SL mechanisms for processing linguistic information in the auditory modality than exposure to either a Catalan–Spanish bilingual or a monolingual Spanish environment.

Now, we turn to the analysis of specificity on different types of foils. This analysis revealed that rejection was more efficient on linguistic material in the auditory modality for both types of foils. In the visual modality, we found the opposite pattern and only for the foils that violated the positional order of elements (fractals and written syllables) within presented triplets. This result is in line with our earlier suggestion that SL is more efficient on language material in the auditory modality and on nonlanguage material in the visual modality.

The foils that maintain the positional order of elements can be rejected based on detecting violations of TPs and/or weak neural activation in memory networks on presentation of foils.[28,52] In addition to these mechanisms based on associative memory and tracking statistical regularities, rejecting foils that violate the positional order of elements also relies on mechanisms related to positional memory. Native language(s) appear to have influenced the efficiency of positional memory mechanisms differently. The disadvantage shown by Catalan bilinguals in rejecting foils that violated the positional order of syllables or sounds may be accounted for by more efficient positional memory mechanisms in Basque bilinguals and Spanish monolinguals than in Catalan bilinguals. This result was unexpected. Further empirical investigation would be required to understand whether this result is robust, a random statistical artifact, or accounted for some confounding differences between participant samples.

In the visual modality, rejection of foils that only violated TPs (foils in which the syllables or fractals maintained their ordinal positions within foil triplets) was not modulated by material type or group, and rejection of foils that also violated the positional order of elements—and thus allowed people to rely on positional memory in rejecting foils—provided a processing advantage for nonlinguistic material. Intriguingly, in the auditory modality, positional memory enhanced rejection efficiency both on the linguistic and nonlinguistic materials, while in the visual modality, it only enhanced rejection efficiency on the nonlinguistic material, highlighting the fact that native languages only affect auditory processing.

SL is a general ability to extract regularities from sensory input and to use these regularities to structure the environment for more efficient processing, to guide behavioral responses, and to distribute cognitive resources between tasks and sources of sensory input. It is supported by different sets of mechanisms in the visual and auditory modalities. However, within these modalities, the sets of mechanisms that operate on specific types of input (e.g., linguistic and nonlinguistic) are similar.[1,22,51] Phylogenetic factors shaped these mechanisms, while ontogenetic factors can only modulate the efficiency of already-existing mechanisms, honing them for a specific type of input. This only occurs in the auditory modality, by providing new targets for selection via exposure to different sets of cues in the sensory input, relevant for modeling the world by each individual.

Limitations of the study and future perspectives

It is important to emphasize that our conclusion is based on a restricted set of linguistic populations (only three populations, for whom Spanish was either the only or one of the native languages), living in relatively rich Western communities with shared cultural values and socioeconomic structures. Further tests are necessary to confirm the effects of acoustic and phonological differences between languages in a bilinguals' inventory on ontogenetic influences on SL abilities. SL ability over the course of ontogenetic development might also be affected by economic stability, lifestyle (e.g., rural versus urban environments, where different flows of information with distinct cues may occur at varying rates), or even cultural values (e.g., extracting information that is relevant for personal success versus community benefits). Controlled experimental studies are necessary if we wish to better understand how language-irrelevant factors might interfere with individual development of SL abilities in the language and nonlanguage domains. Only then we will be able to fully understand how ontogenetic factors modulate the efficiency of SL abilities shaped by phylogenetic influences.

Another interesting direction for future research would be to compare literate and illiterate populations, especially in regard to ontogenetic influences on SL, or to compare populations in which language has transparent orthography (where correspondences between letters and sounds are direct and allow one-to-one mapping, as in most

Romance and Turkic languages) with populations that use opaque orthography (where letter-sound correspondences are arbitrary to a naive reader, as in, e.g., English, Thai, and French). The native languages of all our participants (Spanish, Catalan, and Basque) had transparent orthographies; hence, we would not have been able to detect any effect of orthography depth (the degree to which grapheme-sound mapping deviates from direct one-to-one correspondence), if one indeed exists on the ontogenetic timescale, in these populations.

The current study is limited to SL efficiency measured as recognition accuracy. Additional insights may be gained by considering other aspects of SL efficiency, beyond recognition accuracy. It is possible that ontogenetic influences would have a stronger effect on speed of learning, automaticity, steepness of learning decay, or the transferability of extracted regularities for processing of new input, different from the input from which these regularities were acquired. Reading, for example, is an automatized behavior (it does not require constant conscious monitoring), and automaticity increases as reading proficiency grows. Reading and SL efficiency are correlated;[62–64] hence, as reading skills improve over the course of ontogenetic development, this may affect SL automaticity rather than accuracy. This could be reflected in automatic detection of statistical regularities and extraction of discrete constituents from a continuous stream of syllables, while leaving the processes related to committing these constituents to memory, strengthening memory representations, and retrieving these constituents during the recognition test unaffected. Transparent orthography may promote sound-to-grapheme links, which might facilitate—as far as possible—the transfer of SL skills across modalities. Different writing systems (hieroglyphic, in which logograms represent morphemes or whole words; alphabetic; logographic, including both classical syllabic scripts, in which symbols stand for separate syllables, and consonant-based logographic scripts, in which symbols represent consonants, with vowels—when necessary—represented by modifications to these core consonantal graphemes) might highlight TPs between different elements—syllables, consonantal graphemes, phonemic graphemes, morphemes, or word-like items—thereby modulating the usefulness of such cues and the cognitive load they incur during processing. Familiarity with different writing systems may affect learning speed and decay in particular types of SL experiments. Hence, other measures of SL efficiency, other than recognition accuracy, may also reveal phylogenetic and ontogenetic influences in the visual modality.

It is also important to note that the patterns of results reported here could have been driven by the nature of the stimuli. Some nonlinguistic sounds can be verbalized (e.g., a footstep, a dog howl, and a water-drop), while it is more difficult to verbalize fractals. In this sense, the sound stimuli were more "linguistic" than the fractals, leading to a greater difference between visual stimuli than between auditory stimuli. This emphasizes the challenge in comparing SL across modalities, and the need to compare the linguistic and nonlinguistic domains separately within each modality. Nevertheless, syllabic sequences are still more speech-like (linguistic) than sequences of sounds, providing differences at the scale between linguistic and nonlinguistic extremes in both modalities. Importantly, despite these limitations, we still observed dissociation in SL performance between modalities on different types of material.

Conclusion

In general, our new results replicated the double dissociation observed in the earlier study: SL mechanisms are more efficient on nonlinguistic than linguistic material in the visual modality, but more efficient on linguistic than nonlinguistic material in the auditory modality. This highlights a species-long (phylogenetic) influence on cognitive machinery. Additionally, we observed that life-long (ontogenetic) influences might lead to further adaptations to the cognitive mechanisms underlying SL in the auditory modality. This suggests that in modern humans, SL is more open to adaptation in the auditory than the visual modality. Exposure to more challenging speech environments, where multiple ambient languages use different cues for the same speech processing tasks, leads to the improvement of SL in the auditory modality, but only on speech-like material.

Acknowledgments

Open access funding enabled and organized by Projekt DEAL.

Supporting information

Additional supporting information may be found in the online version of this article.

File S1. ZIP file of sample test tokens and familiarization streams.

File S2. Comparison of statistical learning efficiency between participants with higher and lower scores on rapid naming test in the Catalan language.

Competing interests

The authors declare no competing interests.

References

1. Saffran, J.R. 2003. Statistical language learning. *Curr. Dir. Psychol. Sci.* **12:** 110–114.
2. Erickson, L.C. & E.D. Thiessen. 2015. Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition. *Dev. Rev.* **37:** 66–108.
3. Kahta, S. & R. Schiff. 2019. Deficits in statistical leaning of auditory sequences among adults with dyslexia. *Dyslexia* **25:** 142–157.
4. Karuza, E.A. *et al.* 2013. The neural correlates of statistical learning in a word segmentation task: an fMRI study. *Brain Lang.* **127:** 46–54.
5. Kidd, E. 2012. Implicit statistical learning is directly associated with the acquisition of syntax. *Dev. Psychol.* **48:** 171–184.
6. Kidd, E. & J. Arciuli. 2016. Individual differences in statistical learning predict children's comprehension of syntax. *Child Dev.* **87:** 184–193.
7. Lany, J., A. Shoaib, A. Thompson & K.G. Estes. 2018. Infant statistical-learning ability is related to real-time language processing. *J. Child Lang.* **45:** 368–391.
8. Misyak, J.B. & M.H. Christiansen. 2012. Statistical learning and language: an individual differences study. *Lang. Learn.* **62:** 302–331.
9. Saffran, J.R. 2002. Constraints on statistical language learning. *J. Mem. Lang.* **47:** 172–196.
10. Saffran, J.R. 2018. Statistical learning as a window into developmental disabilities. *J. Neurodev. Disord.* **10:** 35.
11. Siegelman, N. 2020. Statistical learning abilities and their relation to language. *Lang. Linguist. Compass* **14:** e12365.
12. Barakat, B.K., A.R. Seitz & L. Shams. 2013. The effect of statistical learning on internal stimulus representations: predictable items are enhanced even when not predicted. *Cognition* **129:** 205–211.
13. Dotsch, R., R.R. Hassin & A. Todorov. 2017. Statistical learning shapes face evaluation. *Nat. Hum. Behav.* **1:** 1–6.
14. Gebhart, A.L., E.L. Newport & R.N. Aslin. 2009. Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon. Bull. Rev.* **16:** 486–490.
15. Kirkham, N.Z., J.A. Slemmer & S.P. Johnson. 2002. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* **83:** B35–B42.
16. Kikuchi, Y., W. Sedley, T.D. Griffiths & C.I. Petkov. 2018. Evolutionarily conserved neural signatures involved in sequencing predictions and their relevance for language. *Curr. Opin. Behav. Sci.* **21:** 145–153.
17. Milne, A.E., C.I. Petkov & B. Wilson. 2018. Auditory and visual sequence learning in humans and monkeys using an artificial grammar learning paradigm. *Neuroscience* **389:** 104–117.
18. Wilson, B. *et al.* 2013. Auditory artificial grammar learning in macaque and marmoset monkeys. *J. Neurosci.* **33:** 18825–18835.
19. Toro, J.M. & J.B. Trobalón. 2005. Statistical computations over a speech stream in a rodent. *Percept. Psychophys.* **67:** 867–875.
20. Baldwin, D.A. & J.A. Baird. 2001. Discerning intentions in dynamic human action. *Trends Cogn. Sci.* **5:** 171–178.
21. Baldwin, D., A. Andersson, J. Saffran & M. Meyer. 2008. Segmenting dynamic human action via statistical structure. *Cognition* **106:** 1382–1407.
22. Conway, C.M. 2020. How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neurosci. Biobehav. Rev.* **112:** 279–299.
23. Hard, B.M., G. Recchia & B. Tversky. 2011. The shape of action. *J. Exp. Psychol. Gen.* **140:** 586–604.
24. Hard, B.M., M. Meyer & D. Baldwin. 2019. Attention reorganizes as structure is detected in dynamic action. *Mem. Cognit.* **47:** 17–32.
25. Romberg, A.R. & J.R. Saffran. 2010. Statistical learning and language acquisition. *Wiley Interdiscip. Rev. Cogn. Sci.* **1:** 906–914.
26. Alloy, L.B. & N. Tabachnik. 1984. Assessment of covariation by humans and animals: the joint influence of prior expectations and current situational information. *Psychol. Rev.* **91:** 112–149.
27. Scargle, J.D., J.P. Norris, B. Jackson & J. Chiang. 2013. Studies in astronomical time series analysis. VI. Bayesian block representations. *Astrophys. J.* **764:** 167.
28. Ordin, M., L. Polyanskaya & D. Soto. 2020. Neural bases of learning and recognition of statistical regularities. *Ann. N.Y. Acad. Sci.* **1467:** 60–76.
29. Ordin, M., L. Polyanskaya & A.G. Samuel. 2021. An evolutionary account of intermodality differences in statistical learning. *Ann. N.Y. Acad. Sci.* **1486:** 76–89.
30. Elordieta, G. & J. Hualde. 2014. Intonation in Basque. In *Prosodic Typology II*. S.-A. Jun, Ed.: 405–464. Oxford University Press.
31. Hualde, J. 1991. *Basque Phonology*. Routledge.
32. Hualde, J. 1999. Basque accentuation. In *Word Prosodic Systems in the Languages of Europe*. H. Hulst, Ed.: 947–993. Mouton de Gruyter.
33. Hualde, J.I., O. Lujanbio & J. Zubiri. 2010. Goizueta basque. *J. Int. Phon. Assoc.* **40:** 113–127.
34. Larraza, S. 2010. Acquisition of phonology and Spanish–Basque bilinguals' phonological systems. In *Proceedings of Colloque Jeunes Chercheurs en Acquisition du Langage*. 102–106.
35. Frota, S., M. D'Imperio, G. Elordrieta, *et al.* 2007. The phonetics and phonology of intonational phrasing in Romance. In *Segmental and Prosodic Issues in Romance Phonology*.

P. Prieto, J. Mascaró & M.-J. Solé, Eds.: 131–153. John Benjamins.
36. Prieto, P. *et al.* 2015. Intonational phonology of Catalan and its dialectal varieties. In *Intonation in Romance*. S. Frota & P. Prieto. Eds.: 9–62. Oxford University Press.
37. Lleo, C., A. Benet & S. Cortes. 2007. Some current phonological features in the Catalan language. *Catalan Rev*. **21:** 279–300.
38. Prieto, P., M. del Mar Vanrell, L. Astruc, *et al.* 2012. Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Commun*. **54:** 681–702.
39. Wheeler, M. 2005. *The Phonology of Catalan*. Oxford University Press.
40. Frota, S. & P. Prieto. 2015. Intonation in Romance: systemic similarities and differences. In *Intonation in Romance*. S. Frota & P. Prieto, Eds.: 392–418. Oxford University Press.
41. Nespor, M. & I. Vogel. 2007. *Prosodic Phonology*. De Gruyter.
42. Gervain, J., M. Nespor, R. Mazuka, *et al.* 2008. Bootstrapping word order in prelexical infants: a Japanese–Italian cross-linguistic study. *Cogn. Psychol.* **57:** 56–74.
43. Dutoit, T. & H. Leich. 1993. MBR-PSOLA: text-to-speech synthesis based on an MBE re-synthesis of the segments database. *Speech Commun*. **13:** 435–440.
44. Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge University Press.
45. Ladd, R. 2008. *Intonational Phonology*. Cambridge University Press.
46. Kaufman, A.S. & N.L. Kaufman. 2004. *Kaufman Brief Intelligence Test*. American Guidance Service.
47. Scattone, D., D.J. Raggio & W. May. 2012. Brief report: concurrent validity of the Leiter-R and KBIT-2 scales of nonverbal intelligence for children with autism and language impairments. *J. Autism Dev. Disord.* **42:** 2486–2490.
48. Kievit, R.A. *et al.* 2016. A watershed model of individual differences in fluid intelligence. *Neuropsychologia* **91:** 186–198.
49. Gollan, T.H., G.H. Weissberger, E. Runnqvist, *et al.* 2012. Self-ratings of spoken language dominance: a Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism* **15:** 594–615.
50. Conway, C.M. & M.H. Christiansen. 2005. Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* **31:** 24–39.
51. Frost, R., B.C. Armstrong, N. Siegelman & M.H. Christiansen. 2015. Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn. Sci.* **19:** 117–125.
52. Ordin, M., L. Polyanskaya, D. Soto & N. Molinaro. 2020. Electrophysiology of statistical learning: exploring the online learning process and offline learning product. *Eur. J. Neurosci.* https://doi.org/10.1111/ejn.14657.
53. Friston, K. 2005. A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* **360:** 815–836.
54. Badcock, P.B., K.J. Friston, M.J.D. Ramstead, *et al.* 2019. The hierarchically mechanistic mind: an evolutionary systems theory of the human brain, cognition, and behavior. *Cogn. Affect. Behav. Neurosci.* **19:** 1319–1351.
55. Carbonell, J.F. & J. Llisterri. 1999. Catalan. In *In Handbook of the International Phonetic Association. A guide to the use of the International Phonetic Alphabet*. 61–65. Cambridge University Press.
56. Martínez-Celdrán, E., A.M. Fernández-Planas & J. Carrera-Sabaté. 2003. Castilian Spanish. *J. Int. Phon. Assoc.* **33:** 255–259.
57. D'Imperio, M., G. Elordieta, S. Frota, *et al.* 2009. Intonational phrasing in Romance: the role of syntactic and prosodic structure. In *Prosodies. With Special Reference to Iberian Languages*. S. Frota, M. Vigário & M.J. Freitas, Eds.: 59–98. Mouton de Gruyter.
58. Zacharaki, K. & N. Sebastian-Galles. 2021. The ontogeny of early language discrimination: beyond rhythm. *Cognition* https://doi.org/10.1016/j.cognition.2021.104628.
59. Langus, A., J. Mehler & M. Nespor. 2017. Rhythm in language acquisition. *Neurosci. Biobehav. Rev.* **81:** 158–166.
60. Gasparini, L., A. Langus, S. Tsuji & N. Boll-Avetisyan. 2021. Quantifying the role of rhythm in infants' language discrimination abilities: a meta-analysis. *Cognition* https://doi.org/10.1016/j.cognition.2021.104757.
61. Maye, J., J.F. Werker & L.A. Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* **82:** B101–B111.
62. Brice, H., N. Siegelman, M. Van Den Bunt, *et al.* 2021. Individual differences in L2 literacy acquisition. *Stud. Sec. Lang. Acquis.* https://doi.org/10.1017/S0272263121000528.
63. Hung, Y.-H., S.J. Frost & K.R. Pugh. 2018. Domain generality and specificity of statistical learning and its relation with reading ability. In *Reading and Dyslexia*. T. Lachman & T. Weis, Eds.: 33–55. Cham: Springer.
64. Kahta, S. & R. Schiff. 2016. Implicit learning deficits among adults with developmental dyslexia. *Ann. Dyslexia* **66:** 235–250.