amU
Aix Marseille Université

NNT : 0000AIXM0000

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 20 décembre 2024 par

Leonardo PINTO ARATA

The acquisition of multiword sequences in language processing

Discipline
Psychologie

École doctorale
ED356 – Cognition, Langage, Éducation

Laboratoire/Partenaires de recherche
Centre de Recherche en Psychologie et Neurosciences (CRPN)
Laboratoire d'Informatique et des Systèmes (LIS)

Composition du jury

Bénédicte POULIN-CHARRONNAT Directrice de recherche CNRS, Université de Bourgogne	Rapporteuse Présidente du jury
Arnaud DESTREBECQZ Professeur Université Libre de Bruxelles	Rapporteur
Padraic MONAGHAN Professor Lancaster University	Examinateur
Arnaud REY Chargé de recherche CNRS, Aix-Marseille Université	Directeur de thèse
Carlos RAMISCH Maître de conférences Aix-Marseille Université	Co-directeur de thèse

CRPN
CENTRE DE RECHERCHE EN PSYCHOLOGIE & NEUROSCIENCES

cnrs

LIS

Affidavit

I, undersigned, Leonardo Pinto Arata, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific supervision of Arnaud Rey and Carlos Ramisch, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the French national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Place Marseille, date 25 octobre 2024



Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Liste de publications et/ou brevets et participation aux conférences

1) Liste des publications et/ou brevet réalisées dans le cadre du projet de thèse :

    1. Pinto Arata, L., Ordonez Magro, L., Ramisch, C., Grainger, J., & Rey, A. (2024). The dynamics of multiword sequence extraction. *Quarterly Journal of Experimental Psychology*. Doi: 10.1177/17470218241228548
    2. Pinto Arata, L., Tosatto, L., & Rey, A. (2024). Chunking mechanisms in language and other domains. *L'Année Psychologique/Topics in Cognitive Psychology*.
    3. Pinto Arata, L., Ramisch, Carlos., & Rey, A. (Submitted). Language sequence learning and repetition-spacing.
    4. Ordonez Magro, L., Pinto Arata, L., Fagot, J., Grainger, J., & Rey, A. (Submitted). The lifetime of sequential memory traces in the absence of language.

2) Participation aux conférences et écoles d'été au cours de la période de thèse :
    1. Pinto Arata, L., Helal, Y., Ramisch, C., & Rey, A. (2024, novembre 21-24). The effect of repetition spacing on the development of multiword memory traces [Présentation orale], Psychonomic Society 65th Annual Meeting, New York City, États-Unis.
    2. Pinto Arata, L., Ramisch, C., & Rey, A. (2024, juin 5-7). The effect of repetition spacing in multiword sequence learning [Présentation poster & orale], Conference on Interdisciplinary Advances in Statistical Learning, San Sebastian, Espagne.
    3. Pinto Arata, L., Ramisch, C., & Rey, A. (2024, mai 23-25). The effect of repetition spacing in incidental sequence learning [Présentation orale], 11th Implicit Learning Seminar, Marseille, France.
    4. Pinto Arata, L., Ordonez Magro, L., Ramisch, C., Grainger, J., & Rey, A. (2022, juillet 3-4). The dynamics of multiword sequence learning [Présentation poster]. From Associations to Cognition, Marseille, France.
    5. Pinto Arata, L., Ordonez Magro, L., Ramisch, C., Grainger, J., & Rey, A. (2022, juin 1-3). The dynamics of multiword sequence extraction. [Présentation poster], Conference on Interdisciplinary Advances in Statistical Learning, San Sebastian, Espagne.
    6. Pinto Arata, L., Ordonez Magro, L., Ramisch, C., Grainger, J., & Rey, A. (2022, mai 19-20). The implicit encoding of multiword sequences. [Présentation orale], 9th Implicit Learning Seminar, Graz, Autriche.

Résumé

La capacité à comprendre et à produire un nombre apparemment infini d'énoncés nouveaux a longtemps été considérée comme une spécificité du langage humain. Traditionnellement, les mots simples et les règles grammaticales innées sont considérés comme les éléments de base du langage. Cependant, des découvertes récentes ont conduit à adopter une autre perspective, connue sous le nom de théories basées sur l'usage, qui suggère que les règles grammaticales ne sont pas indépendantes de l'utilisation de la langue et que les séquences multi-mots constituent le cœur du langage. Ainsi, la capacité à traiter et à apprendre des séquences multi-mots semble être cruciale pour la compréhension et la production du langage. De nombreuses études expérimentales soutiennent cette proposition, mais les mécanismes qui sous-tendent l'acquisition des séquences multi-mots restent relativement inexplorés.

L'objectif de cette thèse est d'étudier comment les séquences multi-mots sont acquises en temps réel et de décrire les mécanismes qui façonnent leur traitement. Pour ce faire, nous avons étudié l'influence de la répétition, de la nature de la séquence et de l'espacement des répétitions lors de l'apprentissage des séquences. Dans notre première étude, nous avons démontré que les séquences multi-mots sont rapidement acquises après seulement quatre à cinq répétitions, et que chaque mot de la séquence suit une trajectoire d'apprentissage distincte. En outre, l'apprentissage est influencé par le type de séquence, les mots sémantiquement reliés et les expressions idiomatiques sont traitées plus rapidement et mieux mémorisées que les mots non reliés et les pseudo-mots. Dans notre seconde étude, nous avons montré que l'espacement des répétitions affectait le traitement des séquences multi-mots, l'apprentissage diminuant avec l'augmentation de

l'espacement, mais restant observé pour l'espacement le plus large. La troisième étude, conduite chez des babouins, a confirmé ces résultats et a montré que l'apprentissage de séquences repose sur des mécanismes associatifs fondamentaux et que les traces mnésiques des informations séquentielles ont une longue durée de vie. Enfin, dans un chapitre théorique, nous soutenons que les modèles actuels basés sur le chunking ne peuvent pas rendre compte de ces résultats et que les modèles neuro-informatiques récents basés sur l'apprentissage associatif et Hebbien peuvent être plus appropriés pour décrire et comprendre la nature des mécanismes de chunking.

Dans l'ensemble, cette thèse a permis de montrer que les informations séquentielles sont acquises rapidement, développant une trace mémorielle transitoire, et que la répétition joue un rôle clé dans leur consolidation. Les résultats de cette thèse soulèvent de nouvelles questions importantes pour le domaine de l'acquisition du langage et, plus largement, pour le domaine de l'apprentissage statistique et du chunking.

Mots clés : traitement du langage, séquences multi-mots, apprentissage de séquences, apprentissage statistique, chunking

Abstract

The ability to understand and produce a seemingly infinite number of novel utterances has long been considered a hallmark of human language. Traditionally, single words and innate grammatical rules have been thought to be the basic building blocks of language. However, recent findings have led to an alternative perspective, known as the usage-based approach, which suggests that grammatical rules are not independent of language use and that multiword sequences form the core of language. Thus, the ability to process and learn multiword sequences appears to be crucial for language comprehension and production. Numerous experimental studies support this claim, but the mechanisms underlying the acquisition of multiword sequences remain relatively unexplored.

The objective of this thesis is to investigate how multiword sequences are acquired in real-time and to characterise the mechanisms that shape their processing. To this end, we investigated the influence of repetition, the nature of the sequence, and repetition spacing during sequence learning. In our first study, we demonstrated that multiword sequences are rapidly acquired after about four to five repetitions, and that each word within the sequence followed a distinct learning trajectory. Moreover, learning was influenced by the type of sequence, with semantically related words and idioms being processed faster and remembered better than unrelated words and pseudowords. In our second study, we showed that repetition spacing affected the processing of multiword sequences, with learning decreasing with wider spacing, but still observed for the widest spacing. The third study, conducted with baboons, confirmed these findings and showed that sequence learning relies on fundamental associative mechanisms and that memory traces of sequential information are long-lived. Finally, in a theoretical chapter, we argue that

current chunk-based models cannot account for these findings and that recent neurocomputational models based on associative and Hebbian learning may be more appropriate for describing and understanding the nature of chunking mechanisms.

Taken together, this thesis shows that sequential information is acquired rapidly, developing a transient memory trace, and that repetition plays a key role in its consolidation. The findings of this thesis raise important new questions for the field of language acquisition, and for the field of statistical learning and chunking more broadly.

Keywords: language processing, multiword sequences, sequence learning, statistical learning, chunking

Remerciements

Je tiens tout d'abord à remercier **Arnaud Destrebecqz**, **Bénédicte Poulin-Charronnat** et **Padraic Monaghan** d'avoir accepté de faire partie de mon jury de thèse et d'avoir consacré un peu de leur temps à la lecture de ce manuscrit. J'espère que cela vous plaira. C'est un véritable honneur de vous avoir en tant que jury.

Je souhaiterais remercier mon directeur de thèse **Arnaud Rey**, qui m'a encadré tout au long de ce projet doctoral, mais également lors de mon Master. Cher Arnaud, cela fait déjà six ans depuis le jour j'ai décidé d'aller toquer à ta porte pour te demander si je pouvais faire mon TER de M1 avec toi. Je te remercie sincèrement de m'avoir fait confiance depuis le début, et de m'avoir toujours considéré comme un collaborateur, et non comme un simple étudiant. En travaillant avec toi, j'ai appris que la recherche n'est pas faite que d'expériences avec des résultats clairs et nets, comme on voudrait nous le faire croire. Tu m'as toujours incité à « malaxer mes données, à les faire parler », enfin, à tout explorer de fond en comble. Je te remercie pour toutes ces « j'ai 5 minutes » qui se transforment en 2 heures d'échanges passionnants autour d'un sujet, et qui se terminent avec 3 nouvelles manips et 2 nouveaux projets. Enfin, je te remercie pour ton dévouement envers tes étudiants. Je me dis souvent que nous avons une chance inouïe de t'avoir comme directeur. Tu es mon directeur de thèse, mais tu es aussi devenu un bon ami.

Je souhaiterais ensuite remercier mon co-directeur de thèse **Carlos Ramisch**, qui a accepté gentiment de m'encadrer sans nous connaître ni Arnaud ni moi. Cher Carlos, merci beaucoup pour toutes ces années de travail ensemble, pour avoir toujours été à l'écoute et disponible pour suivre mes avancées, y compris corriger ma thèse en un temps record. Grâce à toi, j'ai découvert un tout autre visage des séquences multimots que je n'aurais jamais pu découvrir si je n'étais resté qu'en psycho. Enfin, je te remercie sincèrement pour ton écoute attentive et tes conseils avisés lors des moments difficiles de la thèse. Tu es plus qu'un directeur de thèse, Você é meu amigo.

Je souhaite également exprimer ma gratitude aux personnes qui m'ont accueilli quand je suis arrivé au LPC, maintenant CRPN. Mes chers **Colette** et **Fred**, mes années de master et de thèse n'auraient pas été aussi rigolotes et conviviales sans vous. Vous avez été les piliers de notre labo pendant toutes ces années. Vous serez à jamais dans mon cœur. **Johannes**, merci d'avoir été l'excellent directeur de labo que tu as été, et de nous avoir permis de réaliser la retraite doc/post-doc du labo. Merci de nous avoir toujours fait confiance en tant que représentants des doctorant·e·s. Évidemment, ce passage par le labo n'aurait pas été le même sans tous les chercheurs et chercheuses que j'ai rencontrés. Je remercie particulièrement **Jérémy Béna** pour toutes nos discussions sur les stats, sur la prise de décision et la mémoire, et sur le métier de chercheur. Je souhaite également adresser un remerciement sincère à **Jonathan Grainger**, qui m'a permis de continuer mes recherches alors que je n'avais pas obtenu la bourse de thèse. Jonathan, tu n'as pas hésité

une seconde à me proposer un poste d'ingénieur d'étude, et je t'en suis infiniment reconnaissant.

Évidemment, cette aventure n'aurait pas été la même sans l'ambiance exceptionnelle de notre laboratoire. Un grand merci à tous les doctorants et doctorantes, post-docs et ingénieurs qui ont fait de ce labo mon deuxième chez moi : **Manon**, **Régis**, **Ladislas**, **Laura**, **Julie**, **Camille G**, **Yannick**, **Kim**, **Anaïs**, **Margaux**, **Camille L**, **Anthony**, **Siham**, **Alexey**, **Angélique**, **Lise**, **Clara**, **Justine**, **Jade**, **Ambre**, **Jon**, **Aurélien**, **Rémy**, **JB**, **JP**, **Giorgio**, **Eulalie**, **Cassandre**, **Lucie, Florence, Andréa**, **Yusuf, Chloé, Pauline**.

Dr To, **Laure** merci d'avoir accepté de m'encadrer en M1. Qui aurait pu savoir que c'était un cadeau empoisonné ? En vrai, merci beaucoup pour ton aide tout au long de mon séjour au labo, pour ta bonne humeur, ton écoute et tes encouragements (« bonyour »). **Guillem**, merci beaucoup d'avoir toujours répondu à mes questions stats et de m'avoir appris que la mémoire n'existe pas. Merci également d'être toujours de bonne humeur à chaque fois qu'on se voit. **Thomas**, merci beaucoup pour tous les échanges scientifiques et non scientifiques que nous avons eus lors de tes passages au bureau, et lors de mes passages à Guillestre. C'est toujours un plaisir d'échanger avec toi. Un grand merci aussi d'être mon dealer de photos. **Floor**, thank you for being there since the very beginning, you are such a good friend. Thank you for your kindness, your joy and all the moments we spent together. I hope to see you next time in Amsterdam. **Diya**, thank you for being such a beautiful person. You always know how to encourage me when things get tough. Now

that the thesis is finished, we're back to whisky evenings! Merci **Maud** pour toutes ces années passées ensemble et pour ta bonne humeur constante. Je n'oublierai jamais quand j'ai découvert que tu étais une experte en Python lors de notre écolé de printemps à Nancy. Évidemment, je n'oublierai pas non plus les soirées Mario Kart où tu m'as toujours détruit. **Eole**, je tiens à te remercier pour ta bonne humeur constante et ton énergie débordante. Merci beaucoup de passer tous les matins dans mon bureau pour papoter un peu, pour voir si tout va bien, et surtout de ton soutien ces deux dernières années (merci beaucoup pour toutes les cartes !). Merci aussi de m'avoir permis d'être ton directeur de l'ombre !

Jacob, mon jaja, les mots manquent pour exprimer à quel point je suis heureux d'être ton ami. Qui aurait cru que notre amitié commencerait par le simple fait qu'Arnaud nous ait mis dans le même bureau. Je n'aurais jamais pensé trouver quelqu'un avec qui partager mon « envie de vivre », le même humour et l'amour pour les stats. Tu es *one in a million*. Merci d'être toujours disponible pour m'écouter et me conseiller, partager un cappu, une planche, un verre de vin, un samos, les memes, enfin, tout. En seulement 2 ans, tu es devenu l'un de mes meilleurs amis, un frère, la famille. Je tiens à te remercier de tout mon cœur d'avoir été à mes côtés tout au long du processus de rédaction de cette thèse. Sans toi, cela aurait été bien plus difficile. Il ne reste qu'à planifier notre voyage à Londres amigo mio. Le J c'est le S.

Elisa, it's been one hell of a ride. Cela fait déjà 6 ans que nous avons commencé cette aventure ensemble. Je me souviendrai toujours de nos journées de révisions pour les

examens en M1 et de notre premier voyage à Londres. Ces moments où nous songions à devenir des chercheurs hors-norme, des chevaliers de l'ordre national du Mérite, à reprendre PNPC. Notre histoire est une histoire de symbiose dans laquelle nous nous encourageons mutuellement à aller plus loin. Nous sommes toujours l'un pour l'autre dans les moments difficiles, et surtout, nous sommes là pour partager nos joies et nos accomplissements. Au risque de te copier, tu es ma meilleure amie ma Elisa. Cette thèse n'aurait pas été la même sans ton soutien indéfectible, et ma vie ne serait pas la même si tu n'en faisais pas partie. A nos cappus noisette, à nos moments de doute, à nos conférences incroyables (surtout la cidrerie), à nos apéros aixois, au fait d'être voisins prochainement, à nos conversations sur les stats, à notre futur papier, mais surtout à notre amitié sans faille. Je t'aime fort mon amie.

Je tiens à remercier tous mes amis et amies que j'ai rencontrés en dehors du laboratoire, et qui m'ont accompagné tout au long de ces années en France. Je pense fort à **Alex**, **Adélie**, **Josh**, **Herbert**, **Léa**, **Tony**, **Virgile**, **Valentin**, **Tarek**, **Thomas**, **Yannis**, **Maïté**. Une pensée spéciale pour mes amis et amies du laboratoire de Gent qui m'ont accueilli comme si j'étais l'un des leurs : **Nicolas**, **Evelyne**, **Leslie**, **Victoria**, **Yulong**, **Rita**, **Anh Le**.

Este párrafo se lo dedico a mis amigos chilenos que nunca dudaron de mí, y siempre estuvieron ahí para apoyarme. **Javier**, nunca pensé que seríamos tan buenos amigos y que el día de mi defensa estarías ahí para apoyarme. Pensar que todo comenzó con una partida

de lolaco cuando estábamos en electivo. Desde ese día, nuestra amistad solo ha ido creciendo hasta el punto en que te volviste uno de mis mejores amigos. Yo creo que no tengo palabras para decirte lo importante que eres para mí y lo mucho que aprecio nuestras conversaciones, nuestros cappus en el café del tiempo y nuestras sesiones de malding para quedar bien calvos. Se le quiere amiguito. **Santiago**, amiguito desde la media que estás a mi lado, en las buenas y en las malas. Te estoy agradecido por haberme apoyado siempre a pesar de mis decisiones a veces dudosas. No puedo esperar a vernos de nuevo. **Patricio**, meu pana, usted ha sido mi mejor amigo durante años, desde que nos jugamos unos lolcitos en la media. Cómo olvidar esos Matías pollo, esos burgerazos y esas sesiones de tarreo en mi casa hasta las 6 de la mañana. Tú has sido un pilar fundamental en mi vida, siempre has estado ahí para acompañarme a lo largo de esta aventura. Nunca podré agradecerte lo suficiente por darme tu amistad, por siempre estar ahí esperándome cuando vuelvo a Chile, por siempre escucharme cuando lo necesito. Tu eres el amigo que toda persona desearía tener. Usted no sabe cuánto se le quiere por estos lares mi rey.

Merci à ma famille française, tonton **René**, tía **Paty**, **Caro**, **Thomas**, **René** et **Lucie** d'être toujours à mes côtés et de veiller sur moi. Gracias a mi familia chilena, a mi tía **Viviana**, mi tío **Raimond**, mis primos **Sebastián**, **Vicente** y **Sofía**, mi abuela **Mirella**, mi abuelo **Humberto**, mi tía **Quena**, mi tía **Paulina**, a mi hermana **Tamara** y mis hermanos **Enzo**, **Sebastián** y **Felipe** por siempre estar a mi lado. **Tamara**, ñaña, gracias por todas las anécdotas habidas y por haber, gracias por darme siempre la motivación que necesito

cuando las cosas se ponen difíciles. **Enzo**, nano, gracias por haberme enseñado todo lo que sé del computador, gracias por ser un apoyo incondicional en mi vida. No importa cuántas páginas escriba, nunca serán suficientes para agradecerle a mi madre, **Claudia**, y a mi padre, **Carlos**, por su apoyo incondicional. Gracias por nunca haber cuestionado mis decisiones de vida y por haberme motivado siempre a ser la mejor versión de mí mismo, sin importar lo que quisiese hacer. Gracias por haberme apoyado desde el inicio de esta aventura y por haberme permitido venir a Francia a estudiar, a pesar de las probabilidades bastante aleatorias de lograrlo. Ojalá todas las personas pudiesen tener padres como ustedes. Como diría Claudio Capéo « Je suis riche d'un père et d'une mère ». Ustedes son mi todo, los amo de lo más profundo de mi ser. Esta tesis es por ustedes.

Enfin, ce dernier paragraphe s'adresse à celle qui partage ma vie et qui m'a accompagné tout au long de ce périple. **Eva**, la vie est belle à tes côtés. Merci pour tous ces cafés, ces verres de vin, ces fous rires, ces voyages (London calling, of course), ces concerts, ces câlins réconfortants, ces échanges scientifiques qui me donnent un milliard d'idées. Tu es la chercheuse la plus incroyable que je connaisse, tu es une source d'inspiration et d'admiration pour moi. Tu es la théorie qui donne du sens à mon quotidien. 40 ans de recherche et plus, mon amour. Te amo lo inimaginable.

J'ai débuté cette thèse en terre inconnue,
Je la termine, cœur et âme français.
Merci à la France pour tout.

Table of contents

Affidavit **2**
Liste de publications et/ou brevets et participation aux conférences **3**
Résumé **5**
Abstract **7**
Remerciements **9**
Table of contents **16**
Preface **18**
Chapter 1: Introduction **22**

1. Theory of generative grammar: Language as an innate ability *23*
2. Usage-based approaches: Language acquisition as a combination of usage and domain-general mechanisms *24*
2.1. Multiword sequences: A building block of language *25*
3. Learning mechanisms in usage-based approaches *30*
3.1. Statistical learning *30*
3.2. Chunking *33*
3.2.1. PARSER *35*
3.2.2. TRACX *36*
3.2.3. Chunk-Based Learner *37*
3.2.4. CLASSIC and CLASSIC Utterance Boundary *39*
4. Summary of previous literature and the objectives of this thesis *40*
5. Hebb lexical decision task *42*

Chapter 2: The dynamics of multiword sequence extraction **47**
Chapter 3: Language sequence learning and repetition-spacing **120**
Chapter 4: The lifetime of sequential memory traces in the absence of language **152**
Chapter 5: Chunking mechanisms in language and other domains **177**
Chapter 6: General discussion **228**

6. Summary of findings *229*
6.1. The effect of repetition in language acquisition *231*
6.2. Repetition and predictability effects *233*
6.3. The influence of repetition spacing in sequence learning *234*
6.4. Does repetition always result in the same trajectory for the development of memory traces? *235*
6.5. Limitations of the Hebb lexical decision task *237*
6.6. Future perspectives *238*
7. Conclusion *240*

Bibliographie 242
Table des figures 264

Preface

"*For a large class of cases—though not for all—in which we employ the word 'meaning' it can be defined thus: the meaning of a word is its use in the language*"

Wittgenstein, L. (1953)

Have you ever wondered what your life would be like without language? Could you do all your daily activities without it? Chances are that many of the ordinary things we do would be more difficult, if not impossible. In fact, we use it everyday for a wide range of actions, such as communicating with our peers, expressing our mental states, referring to abstract concepts, talking to ourselves or describing the world around us. Most impressively, we produce language so effortlessly all the time, from simple words to complex sentences that can be combined into meaningful objects such as jokes, tongue twisters, letters, essays, books, and doctoral theses.

The ability to comprehend and produce a seemingly unlimited number of novel utterances has long been considered a hallmark of human language, and the question of how infants acquire such a generative ability has been the subject of extensive empirical research in psychology. To no one's surprise, except perhaps Noam Chomsky's, we are not born with language. Instead, we have to develop the ability to speak and understand our native language during childhood, without any formal instruction. Even as adults, we need to process and learn word sequences we have never heard before in order to adapt, and modify our language to keep pace with our rapidly changing linguistic environment. But how exactly is this achieved?

The goal of my Ph.D. research is to investigate how multiword sequences are acquired and to gain a better understanding of the cognitive constraints that shape their acquisition and development in real-time language processing. The first chapter of this thesis provides an overview of the background literature that served as the foundation for the three experimental studies that will be presented in this manuscript. I start by presenting two competing theoretical approaches to language acquisition, the nativist/generative proposal and the constructivist/usage-based proposal. I then define what a multiword sequence is. Next, I discuss three underlying mechanisms involved in language learning and processing, namely statistical learning and chunking, an associative learning mechanism. I then present relevant computational models of language learning based on these mechanisms. Finally, I present a novel task, the Hebb lexical decision task, that my collaborators and I have designed to study multiword sequence learning.

Chapter 2 presents the first experimental study (Study 1), in which I investigated the extraction and learning dynamics of multiword sequences. Using the Hebb lexical decision task, I studied and compared the learning of multiword sequences composed of: unrelated words, pseudowords, semantically related words, or idioms. Chapter 3 presents the second experimental study (Study 2). It focuses on the influence of repetition-spacing on the acquisition of multiword sequences. In a modified version of the Hebb lexical decision task, participants were presented with a multiword sequence that was repeated every 4th, 7th, 10th, 20th, 30th or 60th trial, depending on the spacing condition. In Chapter 4, I present an experimental study (Study 3) that examined the influence of repetition-spacing on sequence learning in baboons. The aim of this study was to isolate the chunking mechanisms from the influence of individuals' preexisting linguistic knowledge and expectations of regularities during sequence learning.

In Chapter 5, I discuss the results of these three studies, as well as other studies conducted by our research team, and consider their implications for chunking as a domain-general mechanism in sequence learning. I then consider the extent to which chunking can account for the results of the experiments.

Finally, in Chapter 6, I provide a general discussion of the experimental results I have obtained and of the limitations of the present work. I also identify key questions and directions for future research.

Part I

Theoretical background

Chapter 1: Introduction

As the story goes, the first recorded psychological experiment to study the origin of language was conducted by the Egyptian Pharaoh Psammetichus I in around 429 BC. In an attempt to find the root of human language, the original language, he ordered two newborn children to be given to a shepherd with the strict instruction that no one speak to them. The Pharaoh believed that by raising the children in complete isolation and depriving them of any linguistic input, the first word they uttered would reveal the root language. The experiment culminated in one of the children saying *βεκός*, the Phrygian word for bread, leading the Pharaoh to conclude that the original language was Phrygian. Although the Pharaoh's experiment obviously lacked scientific rigour, it clearly illustrates the fact that research into language acquisition has been a subject of great interest since the earliest days of human history.

Over the years, various theories have been put forward to explain how individuals acquire and produce language. However, it was not until the middle of the 20th century that research into language acquisition in the fields of linguistics and psychology reached a turning point with the publication of Noam Chomsky's book "Syntactic Structures" (1957) and his landmark paper "A Review of BF Skinner's Verbal Behavior" (1959), in which he introduced the theory of *generative grammar* to account for language acquisition. The impact of Chomsky's theory was such that it is regarded as one of the pivotal events in the cognitive revolution, shaping the development of the field for almost 50 years. It was not until the late 1980s that a novel theory succeeded in challenging the generative approach. In the two following sections, I will briefly describe both theories of language acquisition.

1. Theory of generative grammar: Language as an innate ability

Generative grammar encompasses the linguistic theories of syntax that postulate the existence of a set of innate abstract grammatical rules that operate over word classes, allowing humans to produce and understand an infinite number of grammatically correct sentences using a finite set of formal rules and vocabulary. In this view, language proficiency results from an innate capacity for language processing shared by all human languages, known as Universal Grammar (UG). UG is defined as the "system of categories, mechanisms and contraints shared by all human languages and considered to be innate" (O'Grady et al., 1997, p. 195). It includes universal principles, which are presumed to be the common aspects and constraints of all natural languages, and parameters, a set of "switches" that can be either turned on or off and are responsible for the grammatical variation between languages. These principles are assumed to be arbitrary and cannot be explained in terms of learning or cognitive constraints (Dąbrowska, 2015).

Over the years, this theory has evolved into different frameworks, such as the *Principles and Parameters Theory* and the *Minimalist Program* (Chomsky, 2014), with the words-and-rules approach being one of the most influential in psychology. Following the core ideas of generative grammar, Pinker (1994, 1998) proposed that not every linguistic construction can be explained in terms of rules. This is particularly the case with the past tense of irregular verbs in English, as well as some idiomatic multiword expressions. Instead, language acquisition would be based on an interplay between a mental lexicon acquired through experience and innate grammatical rules, both of which depend on distinct neural systems (Pinker & Ullman, 2002; Ullman, 2001). Thus, learners memorise single words and morphemes that they encounter, which are then stored in the mental lexicon. These lexical units are combined into complex words, phrases and sentences by applying a set of rules.

Regardless of the framework adopted, the fundamental tenet common to all generative approaches is the so-called *poverty of stimulus*, which posits that natural language lacks the necessary information for learners to acquire it from the input alone (Chomsky, 1980). Interestingly, this idea was challenged around the same time that Pinker proposed the words-and-rules theory. In fact, analysis of corpora of child-directed speech and a growing body of empirical work revealed that the information contained in the input is much richer than previously thought (e.g., Brent & Cartwright, 1996; Cameron-Faulkner et al., 2003; Christiansen et al., 1998; Morgan & Newport, 1981; Redington et al., 1998; Redington & Chater, 1997; Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996). These new findings gave rise to an alternative theoretical perspective on language acquisition known as usage-based approach.

2. Usage-based approaches: Language acquisition as a combination of usage and domain-general mechanisms

In contrast to the generative theory, usage-based approaches reject the existence of any innate linguistic knowledge and argue that language knowledge emerges gradually through language use and generalisation. Grammatical rules are no longer seen as independent of language use, but rather as having been inferred at some point from linguistic experience (Behrens, 2009). Moreover, language acquisition is considered to be driven by children's desire to communicate with others, rather than simply developing according to a hard-wired cognitive blueprint.

According to this view, language acquisition is based on three cognitive processes: entrenchment, categorisation and generalisation. First, learners create memory traces of frequently repeated items and reinforce them with each new repetition of the same item. Secondly, they categorise similar items by comparing stored items with new items.

Finally, once enough items have been stored within the same category, they generalise across them to generate knowledge about the abstract properties of language, including syntax, grammatical categories and semantics (Abbot-Smith & Tomasello, 2006; Behrens, 2009). Contrary to the words-and-rules approach, the stored items are no longer limited to morphemes and single words, but also include constructions of a varying level of complexity (Goldberg, 2006; Ibbotson, 2013).

Constructions are defined as a learned pairing of (written or phonetic) form and function-meaning (Goldberg, 1995; Goldberg & Herbst, 2021). As their definition suggests, constructions contain information about their use at different levels, such as structural properties, semantic properties and pragmatic properties (Goldberg & Suttle, 2010). They can range from simple linguistic forms (e.g. morphemes, single words) to more complex structures such as complex words and multiword sequences (e.g. phrases, formulaic language, idioms, comparative conditional constructions).

Importantly, constructions are not fixed patterns once acquired, but can evolve over time to adapt to language use. Although some constructions may be stored as a whole because of their high frequency or non-compositional nature (Bybee, 2006; Tomasello, 2003). The objective of this thesis is to examine the acquisition of a particular type of construction, multiword sequences. In the following subsection I will provide a definition of a multiword sequence and also present experimental research that supports the claim that multiword sequences are building blocks of language.

2.1. Multiword sequences: A building block of language

Multiword sequences are often defined as a number of consecutive words that are stored and retrieved from memory as a whole (Wray, 2002), thus acting as a single unit and resulting in a processing advantage over a sentence generated on the fly. These

include a wide range of linguistic phenomena such as collocations (e.g., *strong coffee*; *café corsé*; *café fuerte*), binomials (e.g., *black and white*; *noir et blanc*; *blanco y negro*), phrasal verbs (e.g., *pick up*; there seem to be no phrasal verbs in French or Spanish), idioms (e.g., *kick the bucket*; *casser sa pipe*; *estirar la pata*), proverbs (e.g., *walls have ears*; *les murs ont des oreilles*; *las paredes oyen*) and frequent compositional phrases (e.g., *I don't know*; *je ne sais pas*; *no sé*). It is worth noting that this processing advantage may also result from simultaneous access to each component or multiple combinations of the sequence, rather than from access to the sequence as a whole (Wray, 2012).

Before the advent of usage-based approaches, single words and rules were seen as the building blocks of language, and multiword sequences as the product of the combination of these elements. Since then, extensive work on corpus analysis has shown that language is highly repetitive and consists largely of multiword sequences. For instance, Lieven et al. (2003) recorded the speech of a two-year-old infant for six weeks and found that about half of their discourse consisted of multiword sequences, 54% of which had been used previously during the recording. Similarly, Bannard and Matthews (2008) analysed a corpus of one mother's speech to her child and showed that many multiword sequences were as frequent or more frequent than single English words (see Figure 1). More recently, Xu et al. (2023) found that infants begin to produce multiword sequences of varying lengths, from two to five words, as early as 14 months of age, and that their use increases with age. Interestingly, this pattern is not unique to spoken language and has also been observed in written language, where multiword sequences account for approximately 50% of written discourse (e.g., Erman & Warren, 2000; Y. Wang, 2019).

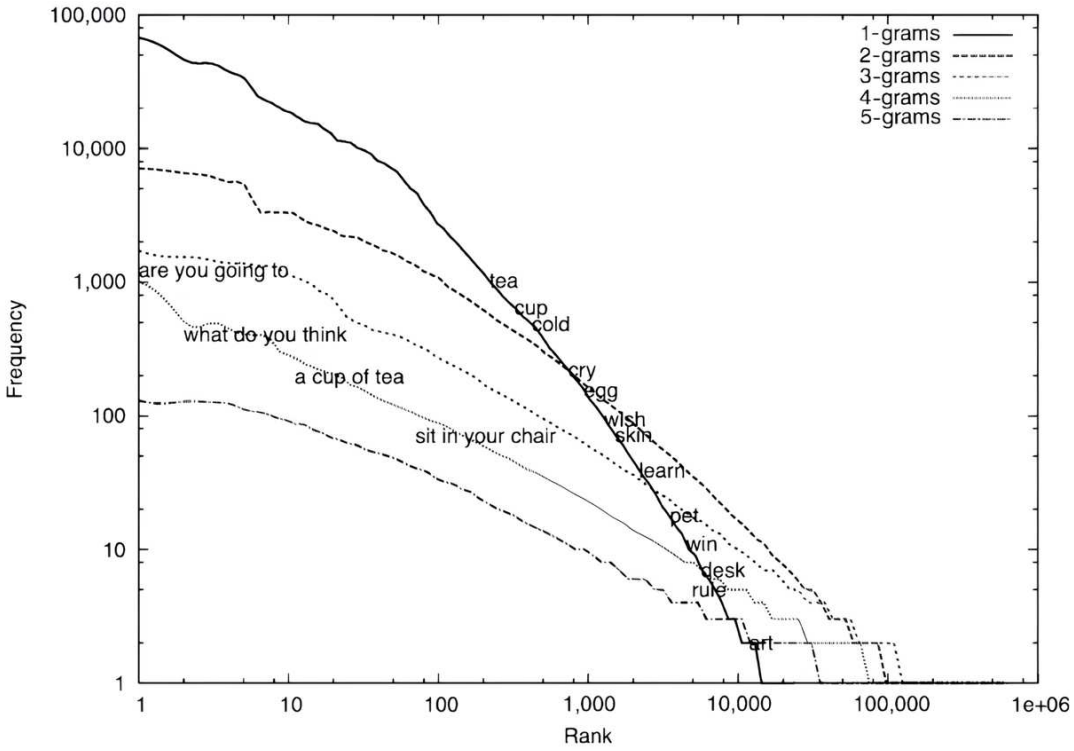


Figure 1. Frequency of n-grams plotted against their rank order, on logarithmic scales. Comparison of four-word sequences and single words. Figure adapted from Bannard and Matthews (2008).

In addition to corpus analyses, a number of developmental studies have provided further evidence that children and adults are sensitive to the statistical properties of multiword sequences and actively use them in language comprehension and production (Abu-Zhaya et al., 2022; Christiansen & Arnon, 2017). For instance, 11- and 12-month-olds are already able to discriminate between frequent (e.g., *clap your hands*) and infrequent multiword sequences (e.g., *take your hands*, Skarabela et al., 2021), whereas 2- and 3-year-olds are more likely to correctly repeat frequent four-word phrases (e.g, *you want to play*) than less frequent ones (e.g., *you want to work*, Bannard & Matthews, 2008; Matthews & Bannard, 2010), and 4-year-olds are better at producing irregular plurals when presented in a familiar context (e.g., *On your – feet*; Arnon & Clark, 2011).

Furthermore, multiword sequences acquired in early childhood have been found to be processed faster in adulthood (Arnon et al., 2017; Elsherif et al., 2020).

Similar to children, adult speakers are faster at recognising and producing frequent multiword sequences (Arnon & Snider, 2010; Janssen & Barber, 2012; Jeong et al., 2021), with high-frequency multiword sequences being better remembered (Tremblay et al., 2011; Tremblay & Baayen, 2010) and shorter to produce in both elicited and spontaneous speech (Arnon & Cohen Priva, 2013, 2014; Siyanova-Chanturia & Janssen, 2018).

Like spoken language, reading behaviour has been shown to be influenced by a variety of multiword sequences. For instance, frequent binomials (e.g., *knife and fork*) are read faster than their reversed forms (e.g., *fork and knife*, Arcara et al., 2012; Siyanova-Chanturia, Conklin, & van Heuven, 2011), idioms (e.g., *at the end of the day* – "ultimately") are read faster and elicit fewer and shorter fixations than non-idiomatic structurally equivalent control phrases (e.g., *at the end of the war*, Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Yaneva et al., 2017), and collocations (e.g., *fatal mistake*) are processed faster than control phrases (e.g., *extreme mistake*, Jiang & Siyanova–Chanturia, 2023; Sonbul, 2015; Vilkaitė-Lozdienė, 2022), even when they are non-adjacent or reversed (Vilkaitė, 2016; Vilkaitė-Lozdienė & Conklin, 2021). Note that sensitivity to multiword sequences during reading does not develop immediately after the acquisition of reading skills. Instead, it seems to emerge around the age of ten (Jiang et al., 2020).

Neuroimaging studies have also provided evidence that the processing of high-frequency multiword sequences differs from that of infrequent or new ones. For example, using MEG, Cappelle et al. (2010) found that listening to existent phrasal verbs (e.g., *rise up*) elicited stronger mismatch negativity than nonexistent ones (e.g., *fall up*), suggesting that existent phrasal verbs are stored and accessed as single lexical units (see also Hanna

et al., 2017, for similar results in EEG). Similarly, using a reading task coupled with EEG, Siyanova-Chanturia et al. (2017) found that frequent binomials (e.g., *knife and fork*) elicited larger P300s and smaller N400s compared to infrequent strongly associated phrases (e.g., *spoon and fork*) and semantic violations (e.g., *theme and fork*), reflecting the preactivation of frequent binomials in memory and their easier processing during reading.

Interestingly, the effect of frequency on multiword sequence processing has been shown to follow a continuum, rather than being an all-or-none phenomenon with either high- or low-frequency multiword sequences. Thus, the magnitude of the processing advantage varies across the continuum from low- to high-frequency multiword sequences (Arnon & Snider, 2010; Hernández et al., 2016; Jiang et al., 2020). Moreover, consistent with usage-based approaches, the processing of multiword sequences is also influenced by other linguistic and distributional properties such as meaningfulness (e.g., Jolsvai et al., 2020), familiarity (e.g., Carrol & Littlemore, 2020), flexibility (e.g., Kyriacou et al., 2020, 2021, 2022), association strength (e.g., Ellis & Ogden, 2017; Li et al., 2021; Yi, 2018; Yi et al., 2017), and predictability (e.g., Chantavarin et al., 2022; Tremblay & Tucker, 2011).

Despite extensive research on the processing of multiword sequences, the question of how they are acquired in real time has been largely overlooked. To date, only two studies addressed this issue in the context of first-language acquisition. In an eye-tracking study, Conklin and Carrol (2020) presented participants with short stories containing existing English binomials in their canonical form (e.g., *brother and sister*), which were seen once, and novel binomials (e.g., *grass and leaves*), which were seen one to five times during the task. Participants were then presented with the existing and novel binomials in reverse order (e.g., *sister and brother*, *leaves and grass*). They found that participants were sensitive to the co-occurrence of the novel binomials, resulting in faster reading times for

the novel binomials as the number of co-occurrences increased. Moreover, the results revealed a processing advantage for forward novel binomials over their reverse forms after only four to five exposures, indicating that participants rapidly detected and encoded the structure of the repeated pattern. Similarly, Pellicer-Sánchez et al. (2022) investigated the learning and processing of novel collocations. Participants were presented with short stories containing existing adjective-noun collocations (e.g., *loud noise*) and novel adjective-pseudoword collocations (e.g., *trobe noise*), which were seen either four or eight times depending on the experimental condition. They found that reading times decreased with exposure, and that reading times for existing and novel collocations were comparable by the fourth exposure.

Taken together, these findings highlight the importance of multiword sequences in language processing and their role as building blocks of language. Most importantly, they highlight a number of key properties that affect their processing, in particular the key role of the frequency of occurrence. In the following section, I will review some of the cognitive mechanisms that have been proposed to underlie the acquisition of multiword sequences, which also serve as the theoretical basis for the work presented in this thesis.

3. Learning mechanisms in usage-based approaches

3.1. Statistical learning

In contrast with the generative assumption that language acquisition can only be explained by innate grammatical knowledge, usage-based approaches assume that language acquisition relies on domain-general mechanisms that allow us to learn the information contained in the input. One of these key mechanisms is the ability to track repeated sequential patterns in the language stream, known as statistical learning (Goldberg & Suttle, 2010; Tomasello, 2003).

Statistical learning refers to the sensitivity and ability of the cognitive system to exploit the statistical properties of the input, such as frequency, variability and probability of co-occurrence, to learn repeated patterns and regularities from the environment (Erickson & Thiessen, 2015). It is regarded as an incidental learning process that occurs by mere exposure (Aslin, 2018), without any intention to learn and without explicit knowledge of what has been learnt, or at least without full conscious awareness (Christiansen, 2019; Cleeremans et al., 1998; Smalle & Möttönen, 2023).

Statistical learning was first described by Saffran, Aslin et al. (1996) to address the problem of language acquisition, namely how infants are able to segment natural speech into words in the absence of clear acoustic cues. In this study, 8-month-old infants were exposed for 2 minutes to a continuous speech stream consisting of four nonsense words of three randomly concatenated syllables (e.g. bidaku-padoti-golabu-tupiro), for which the statistical relationship between the syllables was the only exploitable cue for correct segmentation. Indeed, two syllables forming a word, such as 'bi' and 'da', always appeared together, with the appearance of one predicting the appearance of the other. On the other hand, 'ku' and 'pa' were not fully predictive of each other (see Figure 2). Therefore, the transition probabilities (TPs) of the intra-word syllables (e.g. 'bida') were higher than the TPs of the inter-word syllables (e.g. 'ku-pa'). After exposure, infants successfully discriminated familiar words from novel words (e.g., kudabi, Experiment 1) and from part-words (e.g., dakupa, Experiment 2). These results, together with a replication of these findings in adults (Saffran, Newport, et al., 1996), marked a major turning point in the nature-nurture debate by demonstrating that language can indeed be acquired by relying on environmental cues extracted via domain-general learning mechanisms.

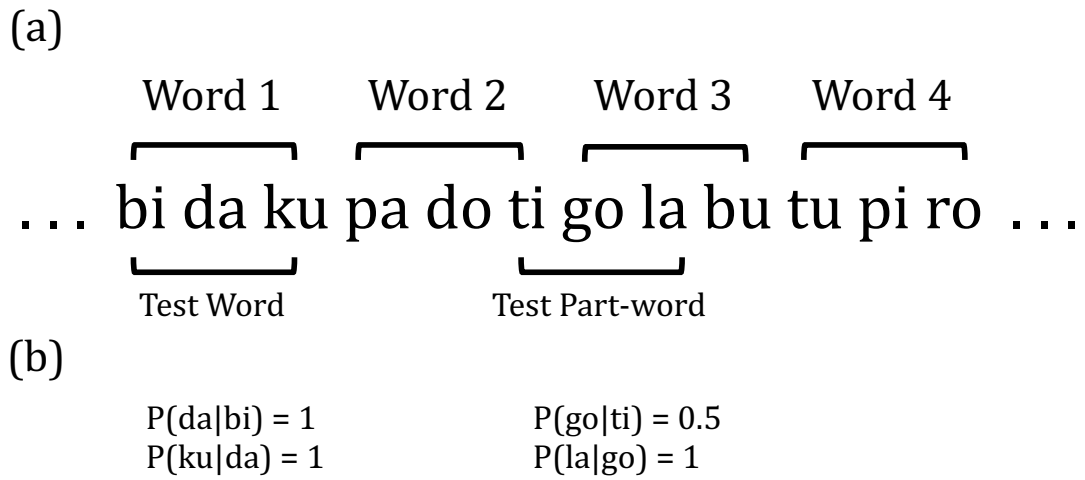


Figure 2. Illustration of the stimuli used by Saffran et al. (1996) to study statistical learning in 8-month-old infants. (a) The list of syllables and three-syllable words. (b) The statistical structure of the words and part-words. Figure adapted from Aslin (2018).

Since the foundational studies by Saffran and colleagues, a substantial body of research has demonstrated that statistical learning is a more powerful mechanism than previously thought, capable of accounting for a wide range of aspects of language acquisition, including word segmentation (e.g., Aslin et al., 1998; Johnson & Tyler, 2010; Pelucchi et al., 2009b, 2009a; Saffran, 2001), phonetical learning (e.g., Maye et al., 2002, 2008; Thiessen & Saffran, 2003, 2007; Weiss et al., 2010), word mapping (e.g., Estes et al., 2007; Smith & Yu, 2008; Suanda et al., 2014; Vlach & DeBrock, 2019; Vouloumanos & Werker, 2009), syntax learning (e.g., Kidd, 2012; Saffran, 2002; Saffran & Wilson, 2003; Thompson & Newport, 2007), and non-adjacent pattern learning (e.g., Frost et al., 2020; Frost & Monaghan, 2016; Gómez, 2002; Newport & Aslin, 2004; Romberg & Saffran, 2010; F. H. Wang et al., 2019). The latter is particularly important for language acquisition, given that the early multiword sequences acquired by children are frames (Tomasello, 2001), a type of construction that has open slots and can contain both adjacent and non-adjacent elements (e.g., *I wanna X*, *I'm X-ing it*, *The Xer*, *the Yer*). Recent behavioural evidence

suggests that the acquisition and production of multiword sequences may also be underpinned by basic statistical learning mechanisms (Isbilen, McCauley, et al., 2022).

Although statistical learning is often referred to as a learning mechanism, it is primarily an umbrella term used to describe the ability to exploit statistical regularities in the environment (Thiessen, 2017). Indeed, the specific underlying mechanisms remain a matter of debate. Among all the proposals that have been put forward, we can distinguish between the TP approach and the chunking approach (Perruchet, 2019). The TP approach assumes that we compute and encode transitional probabilities between successive elements of repeated sequences, whereas the chunking approach assumes that we segment and group repeated sequences into chunks. In this thesis, I will focus on the chunking approach because usage-based approaches have identified it as a crucial process in in the formation and use of multiword sequences (Bybee, 2010; Ibbotson, 2013). Furthermore, converging experimental and computational evidence has also shown that it provides a better fit to human sequence learning data compared to the TP approach (Emerson & Conway, 2023; Isbilen et al., 2020; Perruchet, 2019; Perruchet & Pacton, 2006).

3.2. Chunking

Chunking refers to the process of associating and grouping together co-occurring sequences of items into single processing units called chunks (Gobet et al., 2001). It allows our cognitive system to rapidly compress information to overcome memory limitations and increase short-term memory capacity (Cowan, 2001; Mathy & Feldman, 2012; Miller, 1956), as well as to retrieve and use encoded information faster, thus facilitating several cognitive processes (Christiansen & Chater, 2016; Fonollosa et al., 2015). This process can be either **deliberate** (explicit) or **automatic** (implicit), depending on how and when it

occurs (Gobet et al., 2016). In either case, the main idea behind chunking is that grouping items together into larger units results in the creation of a single memory trace, creating a unified representation of previously separate pieces of information. In this thesis, I will focus on automatic chunking as it is considered to be the main mechanism involved in statistical learning.

Automatic chunking is defined as a non-controlled, unconscious and implicit continuous process in which co-occurring items are grouped together merely by repetition, without any intention to learn. It usually involves processes that occur in long-term memory. Automatic chunking is thought to be essential for language acquisition, as children do not initially store linguistic information using explicit strategies, but rather learn linguistic regularities implicitly (Gobet, 2017). Given the hierarchical nature of language, this type of chunking is also crucial for its processing at different levels. Indeed, our system encodes incoming information into chunks that are progressively reencoded into larger chunks, thereby forming a hierarchical representation of the input consisting of chunks of different sizes and levels of abstraction (Chater et al., 2016; Isbilen & Christiansen, 2018). For instance, at the word-level, letters are grouped together to form larger processing units, such as graphemes and syllable constituents like onsets or rimes (e.g., Brand et al., 2007; LaBerge & Samuels, 1974; Perry, 2023; Rey et al., 1998, 2000; Rey & Schiller, 2005; Ziegler & Goswami, 2005), while at the phrase-level, words are grouped together into multiword sequences and constructions (e.g., Bannard & Matthews, 2008; Bybee, 2010; Carrol & Conklin, 2020; Havron & Arnon, 2020; Pulido, 2021). Recent studies have also shown that chunking can account for the learning and generalisation of non-adjacent regularities (Isbilen et al., 2020; Isbilen, Frost, et al., 2022).

Over the past two decades, the chunking approach has been implemented in several computational models of language acquisition and processing, ranging from single-word

segmentation to multiword sequence acquisition, showing that chunking can successfully account for them. In the following subsection, I will briefly introduce some of the most prominent models in the literature. Although this thesis does not focus on computational modelling, these models will be used to compare their predictions with the results presented in the experimental chapters. A more detailed discussion of the extent to which chunk-based models can simulate current empirical findings will be presented in Chapter 5.

3.2.1. PARSER

PARSER is probably one of the best-known chunk-based models to date. It was developed by Perruchet and Vinter (1998) to account for word segmentation using psychologically plausible processes of memory, attention, and associative learning. Based on the observation that humans tend to segment incoming information into small pieces of variable length, PARSER randomly extracts groups of components (consisting of 1 to 3 components) and temporarily stores them in the percept shaper, a sort of mental lexicon. Each unit is initially assigned a weight. If the same unit is encountered again in the input, its weight increases, otherwise it decreases as a result of decay or interference when processing similar units (see Figure 3). On the one hand, decay eliminates units that do not occur often enough in the input. On the other hand, when the weight of a unit reaches a certain threshold, the model uses it to guide the perception of novel units, allowing it to create longer units from smaller components. As PARSER selects the longest unit to guide perception, shorter units tend to disappear from memory due to interference.

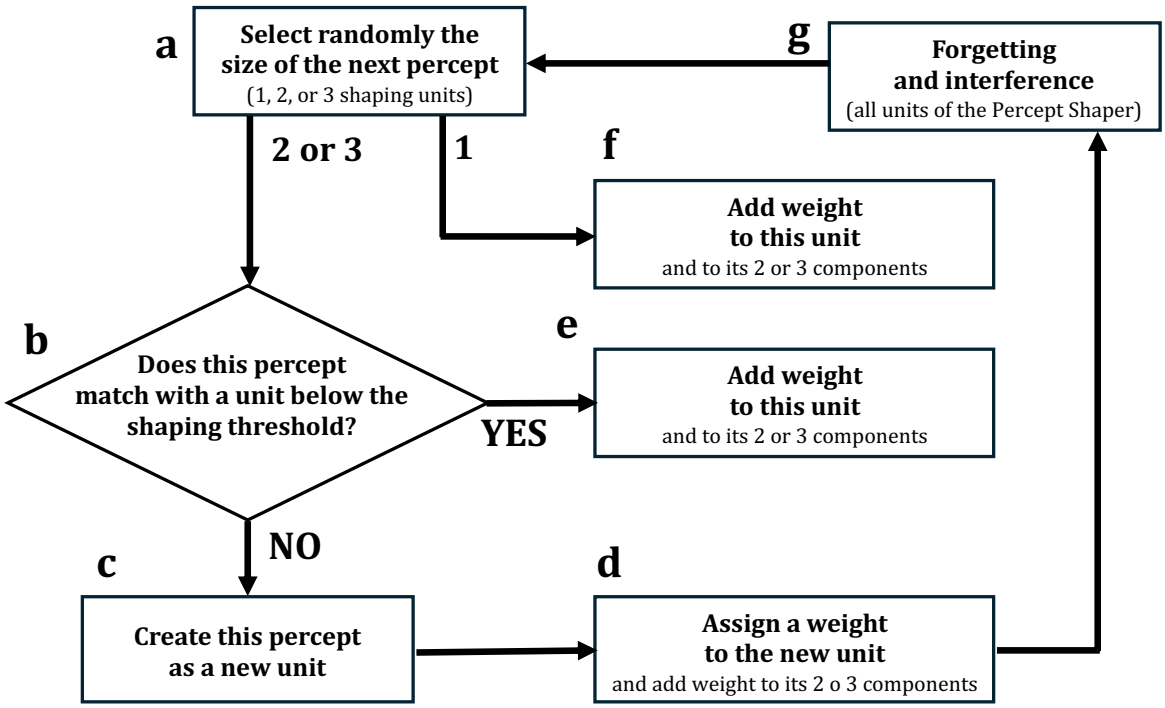


Figure 3. Operations performed by PARSER at each processing step. Figure adapted from Perruchet and Vinter (1998).

3.2.2. TRACX

The TRACX model is a connectionist autoassociator developed by French et al. (2011) that relies on the recognition of previously encountered chunks to account for sequence learning. Unlike PARSER, TRACX has no explicit storage of chunks, instead the information is stored in the synaptic weights of the network. To recognise repeated chunks, the model incorporates a hidden layer that acts as an internal memory that evolves over time as the model encounters new input. In this model, each successive item in a sequence is compared to the previous one, and the combination of the two is compared to the internal memory. Each time the system encounters a pair of items, the output error is reduced by back-propagation. If the error is below a predefined threshold, the pair of items is recognised as a chunk and compared with the next item in the sequence. This process then continues, with each input item being added to the growing chunk until the threshold is

no longer met. However, if the threshold is not met, the system moves on to the next pair of items. More recently, Marechal and French presented TRACX2, a new version of the model in which recognition is no longer an all-or-nothing process, but a gradual process based on output error.

3.2.3. Chunk-Based Learner

A more recent model is the Chunk-Based Learner (CBL), developed by McCauley and Christiansen (2019a). In contrast to PARSER and TRACX, which were developed to model word segmentation, the CBL aims to model language comprehension and production through the discovery and online use of multiword sequences. Indeed, the model assumes that language acquisition occurs in an incremental fashion, on a word-by-word basis, by assembling words into multiword chunks using their distributional information, namely backward transitional probabilities (BTPs). The core unit in this model is therefore the word rather than the syllable.

To simulate language comprehension, the model first compares successive incoming words based on their BTPs. The BTP refers to the probability that an item is preceded by another item. For instance, in the phrase “the dog”, the BTP would reflect how often “the” precedes “dog” in previously encountered language. If the BTPs between them reach or surpass an average BTP, which is dynamically computed based on previously encountered words, the model considers these words to belong to the same chunk and compares them with the next word in the sequence. Conversely, if the BTPs are below the threshold, a boundary is inserted between the words and a chunk consisting of the word or words preceding the boundary is created and stored in the “chunkatory”, a sort of lexicon. These chunks are then used to make online predictions about incoming words and to determine whether they should be chunked together. For instance, when the model

processes a pair of words, it immediately chunks them together if the pair has occurred more than twice before, either as a complete chunk or as part of a longer chunk in the chunkatory. Otherwise, chunking is made based on the BTP between the words (see Figure 4 for an illustration of the process). Importantly, since there is no decay, the words and chunks stored in the chunkatory never disappear.

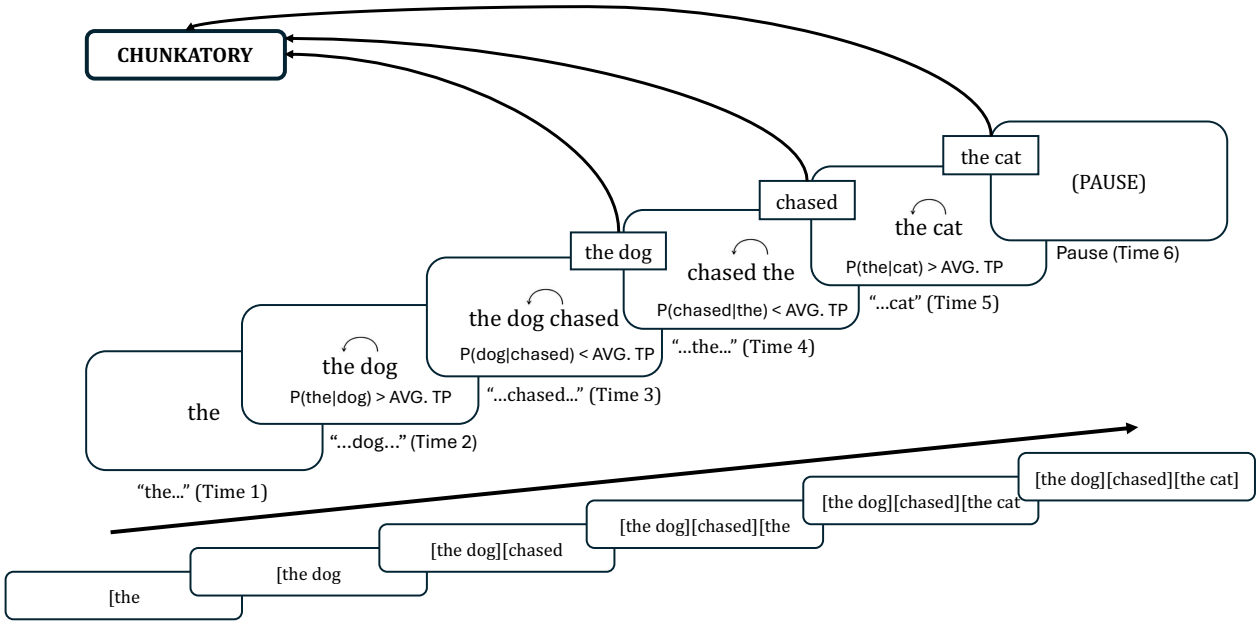


Figure 4. Online processing of the utterance "the dog chased the cat". At each time step, the model calculates the backward transitional probabilities between adjacent words (indicated by the backward arrow's position above the words). If they exceed the average TP threshold, both words or chunks are grouped together. If they are below the running average, they are not grouped and are added to the chunkatory as separate chunks. Figure adapted from McCauley and Christiansen (2019a).

When it comes to simulate language production, each time the model encounters a new utterance, it tries to reproduce it using only its chunkatory and the distributional information it acquired while processing the input. To achieve this, it first compares the parts of the utterance with the chunks present in the chunkatory, starting with the longest chunk in the chunkatory, and then places the corresponding chunks in a bag-of-words in

a random order. The model then attempts to place them in the correct order to reproduce the target utterance by removing from the bag-of-words the chunk with the highest BTP given the start of utterance marker. Once the initial chunk has been placed, the chunk with the highest BTP given the previously produced chunk is removed and placed next. This process is repeated until the bag-of-words is empty. Since language comprehension and production are seen as intertwined processes in this model, the distributional information and the chunkatory are also updated after the reproduction of an utterance.

Finally, it is worth noting that the CBL model is the first usage-based model to have successfully modelled a substantial number of natural language corpora (i.e. 79 single-child corpora for perception and 200 for production evaluation, representing a total of 29 languages). Furthermore, an adapted version of the CBL model has been shown to be able to track and generalise over non-adjacent dependencies to create lexical frames with an open slot (e.g., *on X own* -> *on your own*, *on my own*; McCauley & Christiansen, 2019b)

3.2.4. CLASSIC and CLASSIC Utterance Boundary

CLASSIC (Jones & Rowland, 2017), like the CBL, simulates language acquisition at the multiword level, but uses phonetic information instead of words as input. Thus, the basic units here are phonemes (e.g., aɪ/ n i d/ k ɔ f i/ for "I need coffee"). From this input, the model gradually acquires longer sequences of phonetic information, initially combining co-occurring phonemes into sublexical units (e.g., syllables), then sublexical units into words, and finally, with sufficient exposure, words into multiword sequences. To illustrate this process, I will use the utterance "I need coffee". The first time CLASSIC encounters this utterance, it would encode it into eight chunks (e.g., aɪ/ n i d/ k ɔ f i/), one for each phoneme, and then group them into 6 chunks of adjacent phonemes in memory (i.e., aɪ, ni, id, kɔ, ɔf, fi). Importantly, CLASSIC does not cross word boundaries to chunk

adjacent items unless the chunks in the input correspond to words or multiword sequences. On a second presentation of the same utterance, it would encode it in 5 chunks (e.g., /aɪ, /ni, d, / kɔ, fi/), and then learn new chunks by grouping adjacent items (e.g., nid, kɔfi) proceeding from left to right, using the longest available chunk. In the third presentation, the utterance would be encoded into chunks corresponding to actual words (e.g., /aɪ, /nid, /kɔfi/), allowing CLASSIC to cross boundaries and chunk it into 2 multiword sequences (e.g., aɪ nid, nid kɔfi). Finally, in a fourth iteration, the model would represent the entire utterance as a single multiword chunk (e.g., /aɪ nid kɔfi/). In addition to naturalistic speech acquisition, CLASSIC has also been shown to be able to simulate children's performance on nonword repetition tasks, which are closely related to vocabulary learning (Jones & Rowland, 2017). More recently, Cabiddu et al. (2023) proposed CLASSIC Utterance Boundary, an extended version of CLASSIC that can also perform word segmentation. This is implemented by removing the initial word boundary information in the input and adding utterance boundary information at the beginning and end of the utterance.

4. Summary of previous literature and the objectives of this thesis

In this first chapter, I presented the two main theories that have been put forward to explain language acquisition, focusing specifically on the usage-based theory that serves as the theoretical foundation for this thesis. I then reviewed the experimental evidence supporting the claim that multiword sequences form part of the backbone of human language comprehension and production. Finally, I reviewed the cognitive processes that have been proposed to underpin language acquisition, and by extension multiword sequence acquisition, and presented the main chunk-based models that have been developed to simulate language acquisition.

As I mentioned in the introduction, since the 2000s a growing number of studies have focused on demonstrating that multiword sequences have a special status in memory and play a key role in language acquisition. Much of this literature has focused on investigating the differential processing of high-frequency multiword sequences compared to matched control sequences. In addition, several other studies, primarily using eye-tracking techniques, have focused on what properties of these sequences influence their processing in real time. However, the question of how they are acquired has been largely overlooked. To date, only two studies have addressed this question (e.g., Conklin & Carrol, 2020; Pellicer-Sánchez et al., 2022), focusing exclusively on the acquisition of novel binomials and collocations, and thus leaving aside other types of sequences. Furthermore, in both studies, participants were exposed to a limited number of repetitions of the sequence, making it difficult to know whether their processing would be the same with more exposure (i.e., 5 repetitions for binomials and 8 repetitions for collocations). Therefore, the first objective of this thesis, addressed in Chapter 2, is to investigate the extraction and learning dynamics of different types of multiword sequences, ranging from relatively simple multiword sequences (i.e., unrelated words) to already encoded sequences (i.e., idioms), and to examine how these dynamics evolve with further exposure (up to 45 repetitions).

While the two studies mentioned above have shown that only a few repetitions are necessary to learn a new sequence, less is known about how long this new sequence can survive in memory without being reinforced. According to the associative hypothesis implemented in PARSER, sequences that are not sufficiently reinforced should eventually disappear from memory. To bridge the gap, Chapter 3 focuses on the influence of repetition-spacing (i.e., the distance between each repetition of a repeated sequence) on the acquisition of multiword sequences.

The third objective of this thesis is to isolate and characterise the memory-specific component of chunking during sequence learning. Given the difficulty in isolating explicit and implicit processes in humans, this question is addressed in Chapter 4 by examining the influence of repetition-spacing during sequence learning in baboons. Indeed, humans and non-human primates, are believed to share common associative learning mechanisms necessary for sequence learning (Malassis et al., 2018; Rey et al., 2019, 2022).

Finally, in Chapter 5, I summarise the results of this thesis as well as previous findings on chunking in sequence learning in various domains.

Before proceeding to a presentation of the experimental work (where each study is presented as a stand-alone empirical article, Chapters 2 to 4), in the next section I will briefly introduce a novel task that my collaborators and I designed to study multiword sequence acquisition, and that will be used in the next two chapters of the thesis to study the extraction and acquisition of multiword sequences.

5. Hebb lexical decision task

When studying sequence extraction, there are several ways to assess learning. On the one hand, there are offline measures, which consist in evaluating the representations of repeated patterns that are formed after learning, without looking at the learning process as it unfolds (e.g., 2AFC tasks). This type of measure is useful for assessing whether learning has occurred or for quantifying the overall extent of learning, but it is unsuitable for tracking the learning dynamics of a repeated pattern because it involves other cognitive processes unrelated to sequence extraction, such as decision making (Siegelman et al., 2018). On the other hand, online measures provide information about the learning trajectories of repeated patterns and how their representations are formed during task execution (van Witteloostuijn et al., 2019).

Multiword sequence processing has been studied for the most part using online measures, such as phrase elicitation tasks, self-reading tasks and eye-tracking measures. While these methods offer valuable insights into multiword sequence processing, they usually do not allow us to measure the processing of each item in the sequence accurately, with the exception of eye-tracking. However, eye-trackers require training to use them correctly, and the cost of high-end eye-trackers can be prohibitive (Ivanchenko et al., 2021).

Based on this observation, my collaborators and I propose a novel task that is easy to implement on any computer and that allows us to study in detail how associations between multiword constituents are established over time: the Hebb lexical decision task. This task is an adaptation of the Hebb letter naming task developed by Rey et al. (2020), which combines the Hebb repetition task with a classical regularity repetition task.

The original Hebb repetition task requires participants to recall sequences of digits, with one sequence repeated every third trial. Hebb (1961) found that participants' performance gradually improved on the repeated sequences compared to the non-repeated sequences. In Rey et al.'s (2020) task, participants have to read aloud single letters that are presented one at a time on a computer screen. Unbeknownst to the participants, a triplet of letters (i.e., the Hebb sequence) is repeated with its constituent letters systematically presented in the same order. Similar to the Hebb repetition task, random letters (i.e., fillers) are inserted between two repetitions of the target letter triplets. The learning dynamics of the repeated triplet are tracked by looking at the evolution of response times (RTs) to the letters of the triplet.

To study the dynamics of multiword sequence acquisition, the triplet of letters is replaced by a triplet of words in the Hebb lexical decision task. In this task, participants are asked to decide as quickly as possible whether a sequence of letters displayed on the

screen constitutes a real word or not, using their keyboard. For instance, on French AZERTY keyboards, participants press "M" (for words) or "Q" (for pseudowords). RTs and accuracy are recorded for each word and pseudoword. Each target remains on the screen until the participant responds. The next target then appears immediately after the participant's response (see Figure 5 for an illustration of the task).

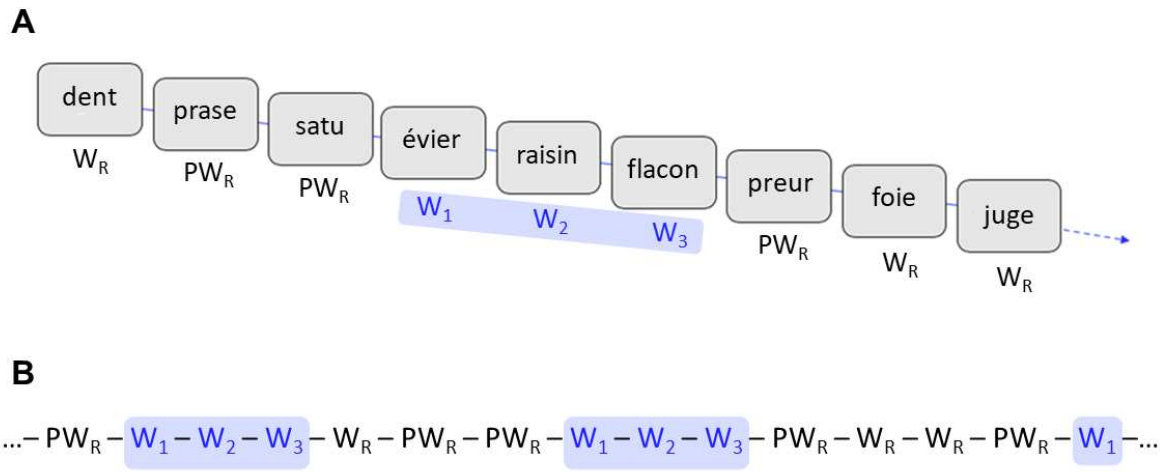


Figure 5. Experimental procedure of the Hebb lexical decision task. Panel A: items are presented one at a time in the centre of the screen. Participants have to classify each string as a word or a pseudoword. A repeated sequence of three words (e.g., W_1: “évier” – sink ; W_2: “raisin” – grape ; W_3: “flacon” – flask), always appearing in the same order, is interspersed with random filler words (W_R) or random filler pseudowords (PW_R). Words in blue belong to the repeated triplet. Panel B: one triplet of words ($W_1W_2W_3$) is repeated several times and a variable number of random words or pseudowords (W_R or PW_R) are presented between two repetitions of the triplet.

We chose to replace the naming task with a lexical decision task because 1) it simplifies data collection and can be easily implemented on crowdsourcing platforms such as Amazon Mechanical Turk or Prolific; 2) subjects must consciously process each item to perform the task correctly; and 3) unlike self-paced or eye-tracking tasks, it is

unambiguous when a subject makes an error during the task, providing us with complementary information about the processing of the sequence.

Part II

Experimental chapters

Chapter 2: The dynamics of multiword sequence extraction

Being able to process multiword sequences is central for both language comprehension and production. Numerous studies support this claim, but less is known about the way multiword sequences are acquired, and more specifically how associations between their constituents are established over time. Here we adapted Rey et al.'s (2020) Hebb naming task into a Hebb lexical decision task to study the dynamics of multiword sequence extraction. Participants had to read letter strings presented on a computer screen and were required to classify them as words or pseudowords. Unknown to the participants, a triplet of words or pseudowords systematically appeared in the same order and random words or pseudowords were inserted between two repetitions of the triplet. We found that RTs for the unpredictable first position in the triplet decreased over repetitions (i.e., indicating the presence of a *repetition* effect) but more slowly and with a different dynamic compared to items appearing at the predictable second and third positions in the repeated triplet (i.e., showing a slightly different *predictability* effect). Implicit and explicit learning also varied as a function of the nature of the triplet (i.e., unrelated words, pseudowords, semantically related words, or idioms). Overall, these results provide new empirical evidence about the dynamics of multiword sequence extraction, and more generally about the role of statistical learning in language acquisition[1].

[1] This experimental chapter is an adapted version of the published manuscript in thesis format. The published version can be found here: Pinto Arata, L., Ordonez Magro, L., Ramisch, C., Grainger, J., & Rey, A. (2024). The dynamics of multiword sequence extraction. Quarterly Journal of Experimental Psychology. Doi: 10.1177/17470218241228548

Introduction

Humans are constantly exposed to and produce an unlimited number of novel utterances and this generative ability has long been considered as a hallmark of human language. For decades, generative linguists have argued that this phenomenon is explained by an innate system of abstract grammatical rules known as the "universal grammar hypothesis" (e.g., Chomsky, 1957). Distinct cognitive abilities supported by different neural systems may allow people to generate complex utterances (Ullman et al., 2005). For example, a mental lexicon including simple linguistic forms (e.g., individual words, morphemes) combined with a mental grammar including combinatorial rules would enable the formation of an infinite number of sentences (Pinker, 1991; Pinker & Ullman, 2002).

More recently, usage-based approaches to language have provided an alternative view to account for the mechanisms involved in language acquisition (e.g., Croft, 2001; Goldberg, 2006; Tomasello, 2003). According to this view, language gradually emerges through the interaction between general cognitive mechanisms and the repeated exposure to concrete items (Ibbotson, 2013). Learners are thought to store incoming utterances and to generate knowledge about the properties of these utterances (e.g., grammatical categories, semantics) by generalising over these stored multiword sequences (Abbot-Smith & Tomasello, 2006).

Over the last two decades, this approach has received multiple computational implementations to illustrate this learning and generalisation process. For instance, Solan et al. (2005) developed an algorithm (ADIOS for automatic distillation of structure) capable of generalising over different kinds of sentences from a given corpus using the statistical information present in the same data. In the same vein, Borensztajn et al. (2009) used an automatic data-oriented parsing procedure to identify the most likely multiword

sequences used in child speech and model the evolution of their abstractness over time. Similarly, Meylan et al. (2017) developed a Bayesian statistical model to study the contribution of language productivity and abstractness to children's linguistic knowledge by focusing on their early capacity to use the determiners "a" and "the" along with a noun.

Whilst these computational modelling studies have successfully captured multiword learning process, the emergence of grammatical knowledge and different developmental patterns more broadly, their reliance on mathematical algorithms and comprehensive corpus analysis undermines their psychological plausibility, as they lack realistic learning mechanisms and memory constraints inherent to the real-time nature of language processing (e.g., Christiansen & Chater, 2016). Chunk-based models, on the other hand, rely on a simple but a powerful mechanism (i.e., associative learning) that can account for both memory constraints and language processing, ranging from single word segmentation (e.g., Perruchet & Vinter, 1998) to multiword sequence acquisition (e.g., Jones & Rowland, 2017). For instance, McCauley and Christiansen (2019) developed a computational model of language perception and production that assumes language acquisition takes place in an incremental fashion, through local shallow processes based on chunking and statistical learning mechanisms. Processing occurs on a word-by-word basis by assembling words into chunks (i.e., sequences of words), rather than via a full syntactic analysis as assumed by generativist theories. Given that language perception and production are thought to be interwoven processes in this model, both are assumed to rely on the same chunks and distributional statistics learnt during language acquisition. Thereby, this model relies on a chunk-by-chunk process instead of whole-sentence optimization. Note that McCauley and Christiansen's model is the first usage-based model having used a large number of natural language corpora (i.e., 79 single-child corpora for perception and 200 for production evaluation, representing a total of 29 languages).

In line with McCauley and Christiansen's (2019) model, numerous studies suggest that language users are sensitive to distributional properties at different levels of the linguistic input, and that statistical learning plays a key role in language acquisition (e.g., Aslin, 2018; Conway et al., 2010; Saffran et al., 1996). For instance, word frequency is known to affect word recognition (e.g., Grainger, 1990) and speech production (e.g., Jescheniak & Levelt, 1994). There is also evidence that linguistic processing is not only affected by word frequency but also by multiword frequency (Ambridge et al., 2015; Carrol & Conklin, 2020). In these studies, a multiword sequence is often defined as a number of consecutive words stored and retrieved from memory as a whole (Wray, 2002), acting as a single unit and resulting in a processing advantage (e.g., "How are you doing?"). It is worth noting, however, that it has also been suggested that this processing advantage could arise from either the simultaneous access to the component parts of a sequence, or from the priming of multiple combinations via the base components, rather than from storing the sequence as a whole (Wray, 2012, p.234).

Many developmental studies have also tested this hypothesis. For instance, Bannard and Matthews (2008) used a sentence repetition task and found that 2- and 3-year-old children are more likely to repeat frequent sentences correctly (e.g., *you want to play*) compared to less frequent ones (e.g., *you want to work*). Arnon and Clark (2011), showed that 4-year-olds are better at producing irregular plurals when presented in a familiar context (e.g., *On your feet*). In the same vein, Janssen and Barber (2012) found multiword frequency effects in adults' production latencies during a task where participants had to name drawings of noun and adjective pairs. Arnon and Snider (2010) also showed that comprehension is affected by multiword frequency. In a grammatical judgement task, adults processed frequent four-word phrases faster than less frequent ones, even when the frequency of the individual final words, bigrams and trigrams were controlled for. It

is worth noting that sensitivity to statistical properties of multiword sequences seems to be present early on. Indeed, it has been shown that eleven- and 12-month-olds can already discriminate frequent multiword sequences from infrequent ones (e.g., *take it off* vs. *shake it of*, Skarabela et al., 2021). Moreover, it has been demonstrated that multiword sequences acquired early in childhood are processed faster in adulthood (Arnon et al., 2017).

Similarly, written language abounds with distributional cues (Arciuli & Simpson, 2012; Snell & Theeuwes, 2020; Treiman et al., 2014). Reading behaviour, for example, has also been shown to be influenced by the frequency and predictability of multiword phrases. For instance, frequent three-word binomial phrases (e.g., *black and white*) are read faster than their reversed forms (i.e., *white and black*) (Siyanova-Chanturia, Conklin, & van Heuven, 2011) and idioms (e.g., *at the end of the day* – 'ultimately') are read faster than non-idiomatic structurally equivalent counterparts (e.g., *at the end of the war*) (Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & Schmitt, 2011).

In the past decades, research has mainly focused on isolated word learning (e.g., Pelucchi et al., 2009; Perruchet & Vinter, 1998; Saffran et al., 1996, 1997), leaving aside the question of how multiword sequences are acquired in real-time. To date, only one study has addressed this issue in the context of first language acquisition. In an eye-tracking study, Conklin and Carrol (2020) presented participants with short stories containing existing English binomials in their canonical form (e.g., boys and girls), which were seen once, and novel binomials (e.g., goats and pigs), which were seen one to five times during the task. Participants were then presented with the existing and novel binomials in reverse (e.g., girls and boys, pigs and goats). They found that participants were sensitive to the co-occurrences of the novel binomials, which translated into faster reading times for the novel binomials as the number of co-occurrences increased. In

addition, the results showed an advantage for forward novel binomials over their reverse forms after only four to five exposures, suggesting that participants very quickly detected and encoded the structure of the repeated pattern (see Sonbul et al., 2022, for a replication in second language acquisition).

Here, we propose to investigate how associations between multiword constituents other than binomials are established over time by using a visual lexical decision task. Based on the assumption that vocabulary acquisition and performance on the Hebb repetition learning paradigm (Hebb, 1961) are subserved by the same processes (Mosse & Jarrold, 2008; Norris et al., 2018; Page & Norris, 2009; Page et al., 2013; Smalle et al., 2016; Szmalec et al., 2009), we used an adaptation of Rey et al.'s (2020) Hebb letter naming task to study the learning dynamics of repeated words triplets.

In the original Hebb repetition task, participants had to recall sequences of digits where one particular sequence was repeated every third trial. Hebb (1961) found that participants' performance gradually improved for the repeated sequences compared to the non-repeated ones. In Rey et al. (2020), participants had to read aloud the names of single letters that were presented one at a time on a computer screen. Unknown to the participants, a triplet of letters (i.e., the Hebb sequence) was repeated with its constituent letters systematically presented in the same order. As in the standard Hebb learning paradigm, random letters (i.e., fillers) were inserted between two repetitions of the critical letter triplets. The extraction dynamics of the repeated triplet was tracked by looking at the evolution of response times (RTs) to the second and third letters of the triplet. RTs for these two letters decreased with repetition as they progressively became predictable when learning occurred. To study the extraction dynamics of multiword sequences in the present experiment, we replaced the triplet of letters used in Rey et al. (2020) by a triplet of words and instead of using a naming task, we used a lexical decision

task hence simplifying online data collection and providing a better proxy for the silent reading that occupies the vast majority of skilled reading behaviour.

The reasons for using the Hebb paradigm to investigate multiword acquisition are two-fold. First, as the Hebb paradigm is an implicit learning measure, it allowed us to study the extraction dynamics of multiword sequences in conditions where participants were not necessarily aware of the repetitions. Indeed, as participants are asked to read words without further instructions, knowledge of patterns of sequences can be attributed to implicit learning through regularity extraction. Second, it allowed us to study the online learning trajectory of multiword sequences rather than solely the "offline" end-product of what has been learned. Indeed, participants' knowledge can be the same at the end of the task (offline knowledge), but their learning trajectories may differ (Siegelman et al., 2017). By using an online learning task, we sought to provide a comprehensive characterization of the process of word-to-word associative learning.

Measuring the evolution of response times for a repeated triplet of items also allowed us to study separately the *repetition* effect from the *predictability* effect. Indeed, because a random number of filler items occurred between two repetitions of the triplet, the first item in the triplet was not predictable and the evolution of RTs for this item can be considered as providing a good estimate of the repetition effect. In contrast, items occurring at Positions 2 and 3 of the triplet benefit from the immediately preceding item that systematically occurs before them and that should help participants anticipating and predicting the next item. Previous studies in sequence learning (e.g., Minier et al., 2016; Rey et al., 2019, 2020, 2022) even reported a stronger predictability effect on the third item of the triplet (i.e., a greater decrease in RTs) due to the richer contextual information provided by the two previous items. This experimental paradigm therefore allowed us to study the differential effect of repetition and predictability on the memory trace of each

item belonging to a repeated triplet and on the processing gains generated by these effects.

Note that the predictability effect is closely linked to chunking mechanisms since it reflects the emergent association between several words that appear repeatedly in a sequence. As previously mentioned, chunking mechanisms are also considered central to several models of sequence learning and language acquisition (e.g., French et al., 2011; Jones & Rowland, 2017; McCauley & Christiansen, 2019; Perruchet & Vinter, 1998; Robinet et al., 2011). However, less is known about the precise dynamics related to the repeated presentation of a sequence of words and empirical evidence is needed to constrain models that assume a central role for chunking mechanisms in the development of language processing skills. The present set of experiments has been designed to provide such empirical evidence about the dynamics of these fundamental associative learning mechanisms.

In the present study, the learning and chunking dynamics of repeated triplets was studied in four Hebb lexical decision experiments. In Experiment 1, the repeated word triplet was composed of three unrelated words. In Experiment 2, the repeated triplet was composed of three pseudowords in order to test if lexicality had an effect on the learning dynamics of the triplet. In Experiment 3, the repeated triplet was composed of three semantically related words in order to test if semantic relatedness would facilitate the development of word associations. In Experiment 4, the repeated triplet corresponded to an existing idiomatic expression to test if the learning trajectory of the repeated triplet would be facilitated by activating the pre-existing long-term memory representation of the triplet. These experiments were conducted remotely by using a platform for online experimentation that has been frequently used in experimental psychology to conduct experiments during the COVID-19 pandemic (e.g., Fournet et al., 2022; Isbilen et al., 2022;

Ordonez Magro et al., 2022). It is worth noting that recent research has shown that JavaScript-based online experiment platforms, such as LabVanced and PsychoJS, allow researchers to collect reliable data that replicate the findings of in-lab studies (e.g., Angele et al., 2022; Mirault et al., 2018).

Experiment 1

Methods

Participants

Forty-two participants (20 females; M_{age} = 24 years, SD = 3) were paid for taking part in the experiment via Prolific (www.prolific.co). All participants reported to be native French speakers, having no history of neurological or language impairment. Before starting the experiment, participants accepted an online informed-consent form. Ethics approval was obtained from the "Comité de Protection des Personnes SUD-EST IV" (17/051).

Given that participants were recruited online, their proficiency in French was measured with the LexTALE language proficiency test (Brysbaert, 2013) before starting the main task. This test consists of a lexical decision task with no time pressure where participants are presented with 84 single-item trials (56 real French words, 28 French-looking pseudowords), and are instructed to decide whether each presented letter sequence is a real French word or not. Their average LexTALE vocabulary score was 86.53% (SD = 5.76). Any participant whose score was below 2.5 standard deviations from the average LexTALE vocabulary score was excluded from the analysis. No participant was excluded based on this criterion. The final dataset consisted of 1890 data points per condition, meeting the 1600 measurements per condition recommendation from Brysbaert and Stevens (2018). A summary of the participants' scores and standard deviations on the LexTALE task for each experiment is provided in Appendix A.

Materials

We adapted Rey et al. (2020)'s naming task into a lexical decision task. The task was composed of 3 blocks of 120 trials, each trial corresponding to the presentation of a single item (word or pseudoword) in the middle of the screen. A set of 66 words and 180 pseudowords were used as items in this experiment. All words were monosyllabic or disyllabic singular nouns. They were composed of four-to-six letters and were selected from the French database Lexique 3.83 (New & Pallier, 2020). Each word of the triplet had a freqfilms2 frequency ranging from 2 to 10 occurrences per million. We decided to use low-frequency words to maximise repetitions effects and increase the chances of revealing any processing differences between positions within the triplet. Indeed, low-frequency words elicit larger repetition effects compared to high-frequency words in lexical decision tasks (e.g., Scarborough et al., 1977). Filler words had a frequency ranging from 10 to 100 occurrences per million. Pseudowords were drawn from the French Lexicon Project (Ferrand et al., 2010). They were monosyllabic or disyllabic and had a length from four to six letters.

A Latin-square design was used such that each word of the triplet appeared in every possible position within the triplet across participants, leading to six possible combinations of the same triplet of words (ABC, ACB, BAC, BCA, CAB, CBA). Seven triplets of words were used and were seen in one of the six possible combinations (for a total of 7*3 = 21 words). Each participant saw one triplet in a specific combination, leading to 7*6 = 42 participants (e.g., Participant 1 saw ABC while Participant 2 saw ACB instead throughout the task). Each triplet appeared 15 times per experimental block (resulting in a total of 45 repetitions across the 3 blocks) and was separated by three to six filler words or filler pseudowords (75 per block). Every block was composed of 60 words (the 15 repeated triplets, i.e., 45 words, and 15 filler words) and 60 pseudowords. Therefore,

there were an equal number of 'yes' and 'no' responses in the experiment (i.e., 180 for each type of response). Among the 66 selected words, 21 served to construct the 7 triplets and 45 served as filler words during the experiment. The set of word triplets and fillers are listed in Appendix B.

In order to obtain more detailed information about participants' explicit knowledge of the task, all participants responded to a short questionnaire after the experiment (similarly to Rey et al., 2020; Tosatto et al., 2022). The first question was: "Did you notice anything particular in this experiment?", in case of a "Yes" response, the follow-up question was "Can you explain what you noticed?" If participants reported noticing the presentation of a repeated sequence of words, they were asked "Can you recall the words in their correct serial order?". If the answer to the first question was "No", the following questions were displayed "Did you notice that a sequence of words was systematically repeated?" and "Can you recall the words in their correct serial order?".

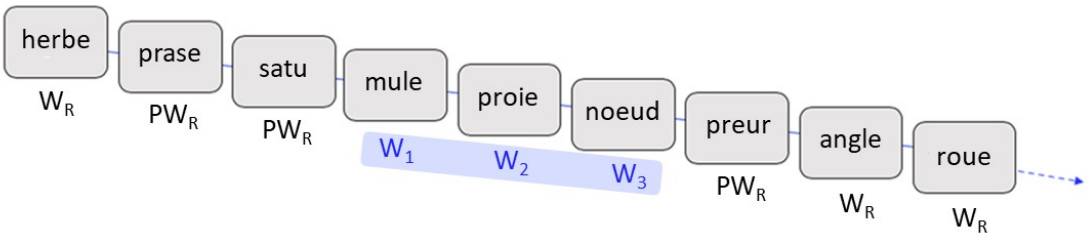
Apparatus

The experiment was implemented in LabVanced, an online experiment builder (Finger et al., 2017) and participants were recruited via the Prolific platform (www.prolific.co). Participants participated via their personal computer and we made sure that the experiment would not work on smartphones or tablets in order to keep the testing conditions as similar as possible across participants. All words and pseudowords were presented in the centre of the computer screen using a 20-point Lato black font on a white background.

Procedure

Before the experiment, written instructions were displayed on the screen. Participants were instructed to decide as fast as possible whether the letter sequence displayed on the screen formed or not a French word. They were required to press "M" (for words) or "Q"

(for pseudowords) on their keyboards (which are at extreme positions on the left and right of French AZERTY keyboards). RTs and accuracy were recorded for each word and pseudoword. Each target stayed on the screen until the participant's response. Subsequently, the next target appeared immediately after the participant's response. To encourage the participants, the number of remaining trials was displayed at the end of each block. The experiment lasted approximately 10 minutes. Figure 1 provides a schematic description of this experimental paradigm.



…– PW_R – W_1 – W_2 – W_3 – W_R – PW_R – PW_R – W_1 – W_2 – W_3 – PW_R – W_R – W_R – PW_R – W_1 –…

Figure 1. Experimental procedure for the Hebb lexical decision paradigm. Upper part: items are presented one at a time at the centre of the computer screen. Participants had to classify each string as a word or a pseudoword. A repeated triplet of three words (e.g., W_1: "mule" – mule ; W_2: "proie" – prey ; W_3: "noeud" – knot) always appearing in the same order was intermixed with random filler words (W_R) or random filler pseudowords (PW_R). Words in blue belong to the repeated triplet. Lower part: one triplet of words ($W_1W_2W_3$) is repeated several times and a variable number of random words or pseudowords (W_R or PW_R) are presented between two repetitions of the triplet.

Results

Only correct trials were analysed (97.06 % of the data), and we excluded RTs exceeding 1500 ms (0.98 % of data) as well as RTs greater than 2.5 standard deviations above a participant's mean per block and for each of the three possible positions within the triplet (2.47 %). The mean RTs and standard deviations computed over the entire sample and for each block are presented in Table 1. Data analysis was performed with the R software (version 4.2.1) using linear mixed-effects models (LMEs) fitted with the lmerTest (version 3.1-3; Kuznetsova et al., 2017) and the lme4 packages (version 1.1-29; Bates et al., 2015). The model included the maximum random structure that allowed convergence (Barr, 2013; Barr et al., 2013), that is, Position (1 to 3), Repetition (1 to 45) and their two-way interaction as fixed effects, participant and Item sets were used as random effects. It is worth noting that Position was coded using repeated contrast coding (i.e., Position 1: -0.7 -0.3; Position 2: 0.3 -0.3; Position 3: 0.3 0.7) in order to perform pairwise comparisons (Schad et al., 2020), and Repetition was mean centred here and in the following analyses. Word length and log-transformed word frequency for each word in the triplet were included as covariates to control for any word-level differences. Given that the distribution of RTs was close to normal and provided good fit (established through visual inspection of QQ plots and histograms), no data transformation was performed prior to the analysis. The results of the model are shown in Table 2.

Table 1

Mean response times (in milliseconds) and standard deviations (in parentheses) for each block and each position in Experiment 1.

	Block 1	Block 2	Block 3
Position 1	617 (100)	572 (95)	581 (120)
Position 2	530 (153)	410 (146)	358 (136)
Position 3	523 (148)	401 (126)	374 (135)

Table 2

Fixed effects of the mixed model for Experiment 1.

Predictors	Estimate	*SE*	95% CI	*p*
(Intercept)	450.75	74.30	[305.12, 596.37]	<.001
Position 2 - 1	-156.54	3.69	[-163.78, -149.31]	<.001
Position 3 - 2	-1.29	3.67	[-8.48, 5.90]	.725
Repetition	-3.95	0.12	[-4.17, -3.72]	<.001
Position 2 - 1 x Repetition	-4.63	0.28	[-5.19, -4.07]	<.001
Position 3 - 2 x Repetition	0.87	0.28	[0.32, 1.43]	.002
Word length	4.38	11.48	[-18.11, 26.88]	0.707
Word frequency (log)	5.58	15.90	[-25.58, 36.75]	0.730

Note. CI: confidence interval; SE: standard error.

We found a significant effect of Repetition with an overall decrease of RTs across the experiment. As predicted, response times for Position 2 were significantly faster than those for Position 1, but they did not differ from Position 3. Moreover, there was a significant negative interaction coefficient for the difference between Position 2 and Position 1, and Repetition, and a significant positive interaction coefficient for the difference between Position 3 and Position 2, and Repetition, indicating that response time differences increased across repetitions. No significant effects were found for word length and word frequency. To investigate where the significant difference between Position 1 compared to Positions 2 and 3 emerges, we ran a series of paired sample *t*-tests on the RTs for Position 1 and the average RTs for Positions 2 and 3 on each repetition of the triplet. We found that a significant difference emerged on the fifth trial, $t(38) = 5.26$, Bonferroni-adjusted $p < .001$.

To get a clearer picture of the learning dynamics for each position in the triplet of words, Figure 2 represents the evolution of the mean response times for each position in the triplet and for the successive 45 repetitions of the triplet. Given that linear regression only captures the overall change of Position across repetitions, we conducted a broken-

stick linear regression, using the segmented package (version 1.6-0; Muggeo, 2008), in order to account for the evolution of the learning pattern across the task. In broken-stick regression, multiple linear regressions are fitted and connected at certain estimated values referred as breakpoints. At the breakpoint the relationship between the variables changes to model non-linear relationships between two variables. Thus, each position was regressed onto Repetition separately. To estimate the number of breakpoints for each position, a broken-stick regression model was built incrementally (i.e., we added a breakpoint estimate to each successive model). For each model, an initial guess for the breakpoint was provided, and then the optimal breakpoints were calculated by the model using an iterative fitting procedure with the default package parametrization (see Muggeo, 2008, for technical details). We compared each new model with the previous one (based on chi-squared analysis) and selected the most parsimonious as the final model. For Position 1, the analysis revealed a breakpoint at repetition 18.46, 95% CI [14.26, 22.67], with RTs decreasing from repetitions 1 to 18.46, b = -4.22, 95% CI [-5.70, -0.29], followed by a slow increase, b = 0.52, 95% CI [-0.29, 1.33]. For Position 2, we estimated two breakpoints at repetitions 5.35, 95% CI [3.54, 7.16] and 19.81, 95% CI [13.63, 25.98], with RTs rapidly decreasing from repetitions 1 to 5.35, b = -32.40, 95% CI [-46.24, -18.57], continuing to decrease, but at a slower rate, from repetitions 5.35 to 19.81, b = -7.92, 95% CI [-10.78, -5.07], followed by a slower decrease until the end of the task, b = -3.20, 95% CI [-4.34, -2.07]. For Position 3, we also estimated two breakpoints at repetitions 5.88, 95% CI [3.52, 8.24] and 19.54, 95% CI [15.36, 23.71], with a fast decrease in RTs from repetitions 1 to 5.88, b = -27.92, 95% CI [-41.01, -14.85], continuing to decrease at a slower rate from repetitions 5.88 to 19.54, b = -8.18, 95% CI [-10.86, -5.50], and with an even slower decrease from repetition 19.54 until the end of the experiment, b = -1.72, 95% CI [-2.78, -0.65].

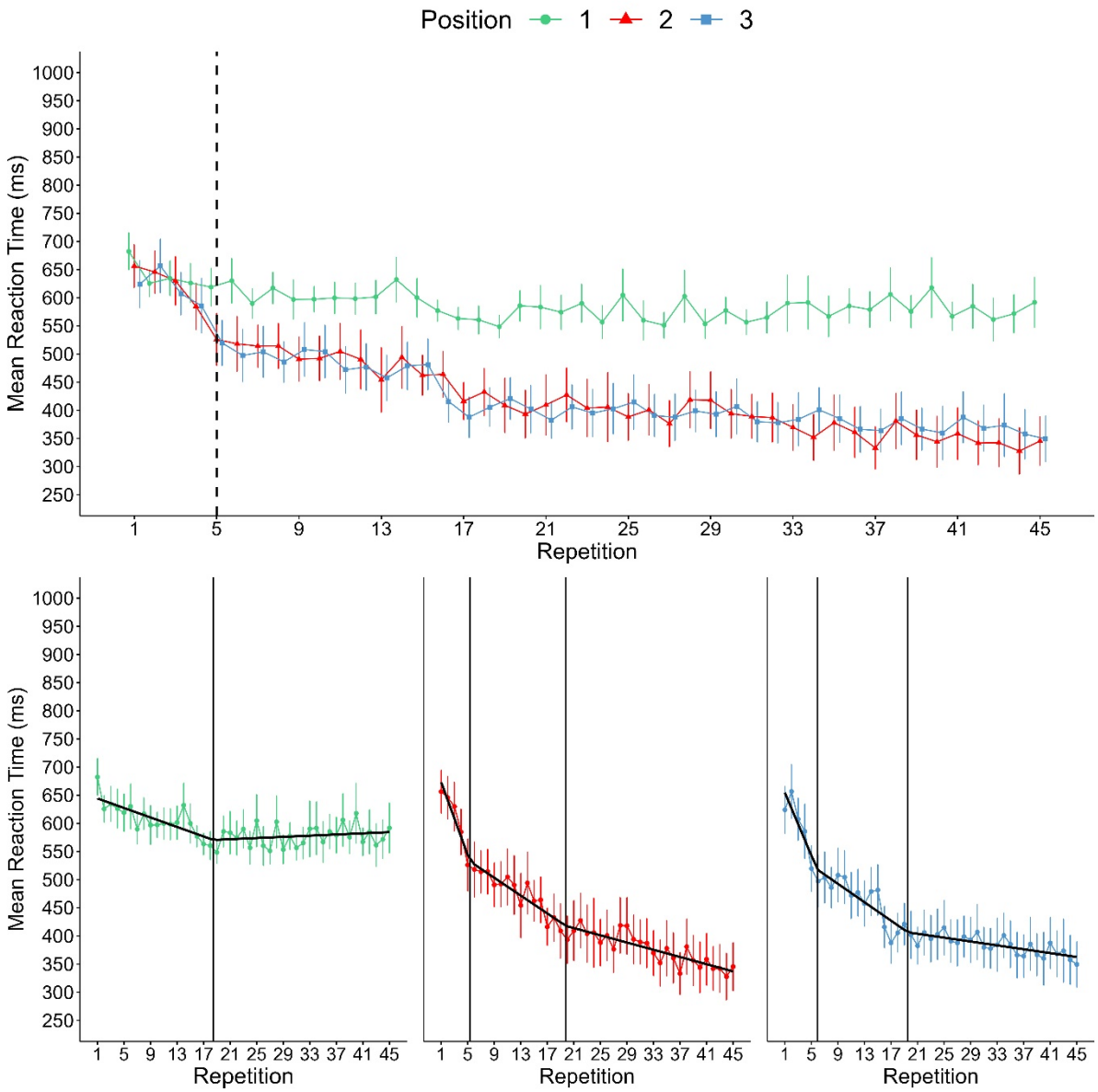


Figure 2. Upper panel: mean response times in Experiment 1 as a function of word position and number of repetitions of the triplet. The vertical dashed line indicates the first repetition at which there was a significant difference between Position 1 vs. Positions 2 and 3. Error bars indicate 95% confidence intervals. Lower panel: results from the broken-stick regressions for each Position in the triplet. Vertical bars indicate the breakpoints.

Questionnaire

Forty-one of the 42 participants reported noticing a recurrent word sequence; 16 were able to recall the whole triplet, 12 correctly recalled one sub-sequence (words 1 and 2 or words 2 and 3), four could recall non-adjacent words (words 1 and 3), seven only recalled one word, and the three remaining participants did not recall any word.

Discussion

As expected, the results from Experiment 1 showed faster RTs for predictable words (i.e., words 2 and 3) within the repeated triplet, and the difference between unpredictable (word 1) and predictable items increased as the task progressed. Furthermore, this difference between unpredictable and predictable items emerges early on, around the fifth repetition of the triplet. The analysis of the mean response times over the 45 repetitions of the triplet further indicated that learning occurred also for words appearing in Position 1 of the triplet. Although unpredictable, these words were repeated and their processing was facilitated by this repetition. The broken-stick regression analysis suggested that learning occurred during the first 18 repetitions and subsequently reached a plateau performance. While the mean RT for the first occurrence of these words was 682 ms, the mean RT was 561 ms after 18 repetitions, and 592 ms at the 45th repetition, indicating a processing speed up of 90 ms between the first and last occurrence of the word. These data therefore provide an estimate of the dynamics of the repetition effect for words that are not predictable.

In contrast, RTs for predictable words (i.e., on Positions 2 and 3) followed a totally different dynamic. According to the broken-stick regression analysis, they indeed decreased very rapidly during the first 5 repetitions (640 ms at the first repetition, and 523 ms at the 5th repetition - RTs are averaged over Position 2 and 3) and the decrease was slower between repetition 5 and 18 (419 ms at the 18th repetition). After the 18th

repetition, RTs continued to decrease but at an even slower rate (347 ms at the 45th repetition). Clearly, compared to the results obtained for words at Position 1 of the repeated triplet, we found that the predictability effect was much larger than the repetition effect and followed different learning dynamics. For example, for the 3rd position of the triplet, the mean response times were 624 ms for the first occurrence of the word and 349 ms for the 45th repetition, resulting in a processing gain of 275 ms between the first and last occurrence of these words.

Interestingly, there was no evidence for an advantage of the third over the second word in the triplet, contrary to what was observed by previous studies. Indeed, prior findings indicated faster RTs for the final stimulus in a repeated triplet, as it benefits from the cumulative information provided by the two preceding stimuli (e.g., Minier et al., 2016; Rey et al., 2019, 2020, 2022). Regarding our study, although the words clearly benefited from immediate contextual information (i.e., the preceding word in the triplet that systematically appeared before them), we did not observe any additional predictability effect regarding the final word of the triplet when the context was richer (i.e., words in Position 3 of the triplet benefit from the contextual information provided by words in Position 1 and 2). This intriguing result likely reflects some limitations of associative and Hebbian learning mechanisms due to the specific time-scale of the present experimental paradigm. We will return to this issue in the general discussion.

Despite a clear decrease in RTs for the predictable positions in the triplet, indicating that learning of this repeated sequence occurred, most participants were unable to correctly recall the whole triplet, even though most of them noticed the presence of a repeated sequence. This result suggests that part of the triplet learning was explicit but that most of the learning was probably implicit. Participants did not have to explicitly

encode the triplet repetition to anticipate the occurrence of words appearing on predictable positions.

In contrast to Experiment 1, which was conducted with triplets of unrelated words, Experiment 2 was conducted with triplets of pseudowords. We decided to use pseudowords because tasks consisting of the repetition and encoding of pseudoword sequences have been shown to mimic novel word learning (Norris et al., 2018; Schimke et al., 2021). Indeed, whereas words are likely to have long-term memory representations, pseudowords cannot benefit from such representations as they have not yet been encountered by participants. It is worth noting that the Hebb paradigm has also been described as a laboratory analogue of novel word learning (Szmalec et al., 2009, 2012). Therefore, studying triplets of pseudowords will allow us to compare the learning dynamics of completely novel multiword sequences with those obtained for already known words in Experiment 1.

Experiment 2

Methods

Participants

Forty-six participants (22 females; M_{age} = 25 years, SD = 3) were recruited from Prolific (www.prolific.co) for the experiment. All participants indicated that French was their native language and declared no neurological or language impairment. Four participants were excluded from the analyses due to chance-level performance on the main task.

As in Experiment 1, participants' French proficiency was measured with the LexTALE test (Brysbaert, 2013). Participants' average scores were 85.13% (SD = 7.08). No participant was excluded from the analysis. The final number of participants was 42, which corresponds to a dataset of 1890 data points per condition.

Materials

In contrast to the previous experiment, here the target triplets were composed of pseudowords whereas the words served only as fillers items. We selected 180 words from the French database Lexique 3.83 (New & Pallier, 2020). All words were monosyllabic or disyllabic singular nouns and had a length from four to six letters. Their freqfilms2 frequency was between 10 and 100 occurrences per million. A set of 66 pseudowords was selected from the French Lexicon Project (Ferrand et al., 2010). Twenty-one were drawn therefrom to construct triplets and the remaining 45 were used as filler pseudowords. All pseudowords were four-to-six letter long and monosyllabic.

Seven triplets were generated and counterbalanced across participants using a Latin-squared design. Every triplet repetition of pseudowords (15 per block) was always separated by three to six filler words or filler pseudowords (75 per block). As in Experiment 1, each block was composed of 60 words and 60 pseudowords. There were an equal number of 'yes' and 'no' responses in the experiment (i.e., 180 for each type of response). The sets of pseudoword triplets and fillers are listed in Appendix C.

Apparatus and procedure

The apparatus and procedure were identical to the one used in Experiment 1.

Results

As the target triplets were made up of pseudowords, only correct "no" responses were analysed (95.87% of the data), and RTs exceeding 1500 ms (1.43% of data), as well as RTs beyond 2.5 standard deviations from a participant's mean per block and for each of the three possible positions within the triplet (2.01%) were excluded. Means and standard deviations per block are shown in Table 3. The linear mixed model we fitted included the maximum random effect structure allowing convergence (Barr, 2013; Barr et al., 2013). This model included Position, Repetition, and the interaction term as fixed effects. Item

and participant were used as crossed random effects, with by-participant random slopes for Position. The results of the mixed model are summarised in Table 4. Figure 3 provides the evolution of mean response times for each position in the triplet and for the 45 repetitions of the triplet.

Table 3

Mean response times (in milliseconds) and standard deviations (in parentheses) for each block in Experiment 2.

	Block 1	Block 2	Block 3
Position 1	719 (134)	670 (125)	641 (112)
Position 2	633 (220)	457 (177)	425 (187)
Position 3	595 (190)	439 (161)	401 (169)

Table 4

Fixed effects of the mixed model for Experiment 2

Predictors	Estimate	*SE*	95% CI	*p*
(Intercept)	554.71	12.50	[530.21, 579.21]	<.001
Position 2 - 1	-171.02	18.51	[-207.30, -134.73]	<.001
Position 3 - 2	-28.43	7.81	[-43.73, -13.13]	<.001
Repetition	-5.39	0.15	[-5.68, -5.10]	<.001
Position 2 - 1 x Repetition	-4.85	0.37	[-5.58, -4.12]	<.001
Position 3 - 2 x Repetition	0.91	0.36	[0.21, 1.62]	.01

Note. CI: confidence interval; SE: standard error.

Results indicated a significant effect of Repetition and faster RTs for pseudowords in Position 2 compared to those in Position 1, as well as for Position 3 compared to Position 2. Moreover, there was a significant negative interaction coefficient for the difference between Position 2 and Position 1, and Repetition, and a significant positive interaction coefficient for the difference between Position 3 and Position 2, and Repetition. Similarly to Experiment 1, paired sample *t*-tests comparisons showed a significant difference

between Position 1 compared to Positions 2 and 3 on the sixth trial, $t(36) = 3.12$, Bonferroni-adjusted $p = .021$.

As for the first experiment, we conducted a broken-stick regression to study the evolution of the learning pattern of Position across the task. The analysis revealed a breakpoint at repetition 5.18, 95% CI [3.58, 6.78] for Position 1, with RTs decreasing from repetitions 1 to 5.18, b = -23.81, 95% CI [-36.78, -10.86], followed by a slower decreasing rate, b = -1.82, 95% CI [-2.38, -1.26]. For Position 2, two breakpoints were estimated at repetitions 7.44, 95% CI [5.59, 9.29] and 22.00, 95% CI [17.56, 26.44], with RTs rapidly decreasing from repetitions 1 to 7.44, b = -39.33, 95% CI [-50.61, -28.05], continuing to decrease, but at a slower rate from repetitions 7.44 to 22.00, b = -10.16, 95% CI [-14.00, -6.32], followed by a slower decrease until the last repetition, b = -1.16, 95% CI [-2.87, 0.56]. Regarding Position 3, we estimated two breakpoints at repetitions 6.07, 95% CI [4.40, 7.74] and 18.37, 95% CI [14.74, 21.99], with RTs decreasing fast from repetitions 1 to 6.07, b = -41.73, 95% CI [-54.71, -28.74], steadily decreasing at a slower rate from repetitions 6.07 to 18.37, b = -11.66, 95% CI [-16.00, -7.33], followed by a slower decrease until the end of the task, b = -1.86, 95% CI [-3.13, -0.59].

Given that usage-based theories postulate that novel items become lexicalised when they are encountered sufficiently often (e.g., Bybee, 2006; Zang et al., 2023), one might expect that after enough repetitions participants would begin to consider the target pseudowords to be almost as real words, resulting in more false “yes” judgments as the experiment progressed. Therefore, we conducted an additional analysis using a generalised (logistic) linear mixed model to compare the mean accuracy between positions across blocks (see Figure 4). The model was fitted with Position and Block, and the interaction term as fixed effects. The maximal random effects structure that converged was one that included by-participant and by-item random intercepts. To explore

differences between positions within each block, we used the R package emmeans (Lenth, 2023). Helmert contrasts were used to compare Position 1 to both Positions 2 and 3, simultaneously, and to compare Position 2 to Position 3. The results of the contrasts are summarised in Table 5. The analysis showed that, systematically across the three blocks, participants made more false "yes" judgments for pseudowords in Position 1 than for those in Positions 2 and 3. In addition, in Block 3, participants made more false "yes" judgments for pseudowords in Position 2 compared to those in Position 3. Finally, false "yes" judgments for pseudowords in Position 1 increased across the blocks, in contrast to those in Positions 2 and 3.

Table 5

Summary of Helmert contrasts between positions across blocks for Experiment 2

	Block 1			Block 2			Block 3		
Predictors	*b*	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
P1 vs P2-P3	-0.76	0.22	<.001	-2.15	0.25	<.001	-2.48	0.25	<.001
P2 vs P3	-0.24	0.30	.42	-0.22	0.43	.60	-1.40	0.44	.002

Note. P: Position; SE: standard error.

Questionnaire

Thirty-nine participants reported noticing a recurrent pseudoword sequence; 12 were able to recall the whole triplet, one could recall one subsequence (words 2 and 3), eight correctly recalled non-adjacent pseudowords (words 1 and 3), eight only recalled one pseudoword, and the 13 remaining could not recall any pseudoword.

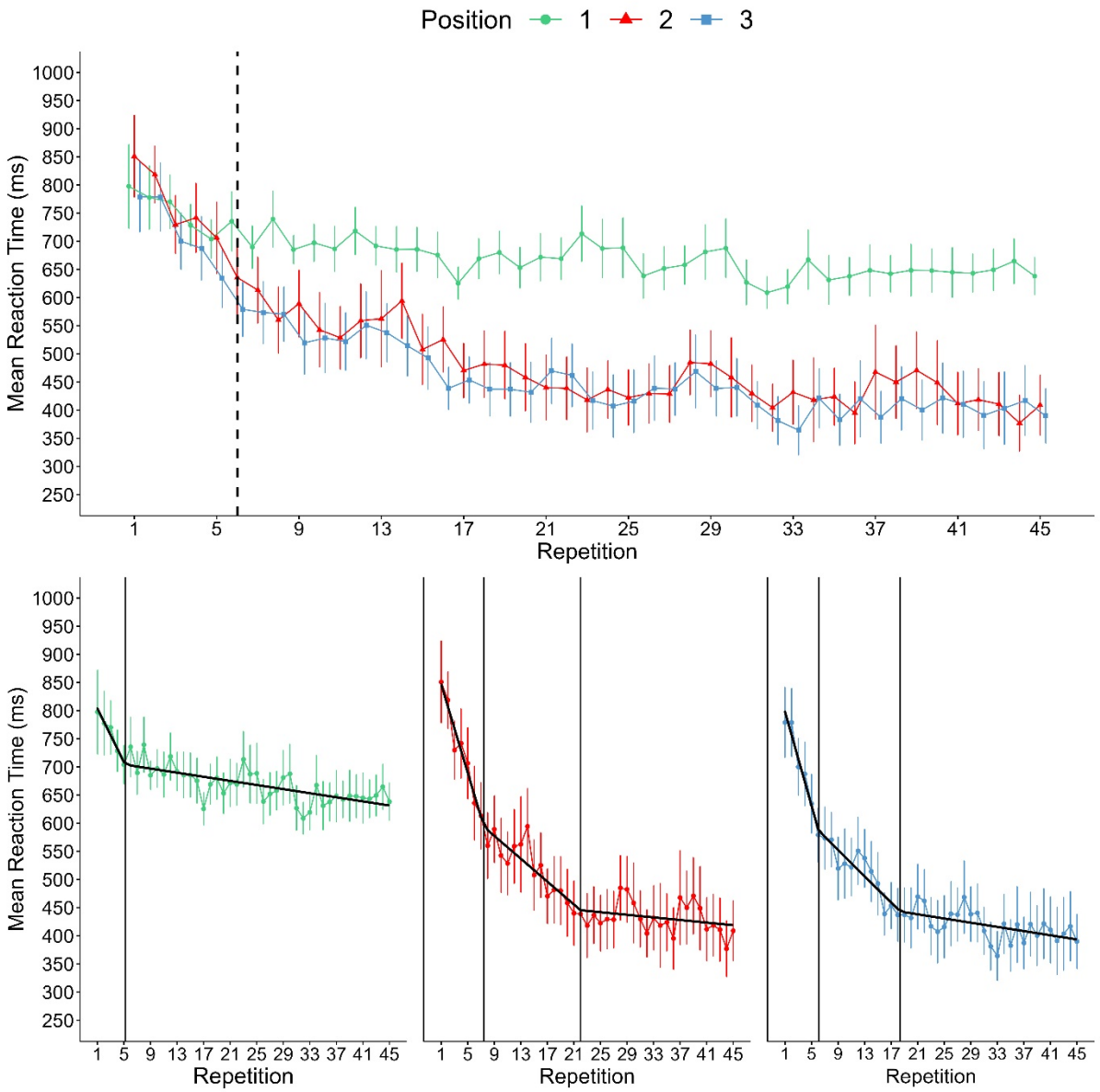


Figure 3. Upper panel: Mean response times in Experiment 2 as a function of pseudoword position in the repeated triplet and number of repetitions. The vertical dashed line indicates the first repetition at which there was a significant difference between Position 1 vs. Positions 2 and 3. Error bars indicate 95% confidence intervals. Lower panel: Results from the broken-stick regressions for each Position in the triplet. Vertical bars indicate the breakpoints.

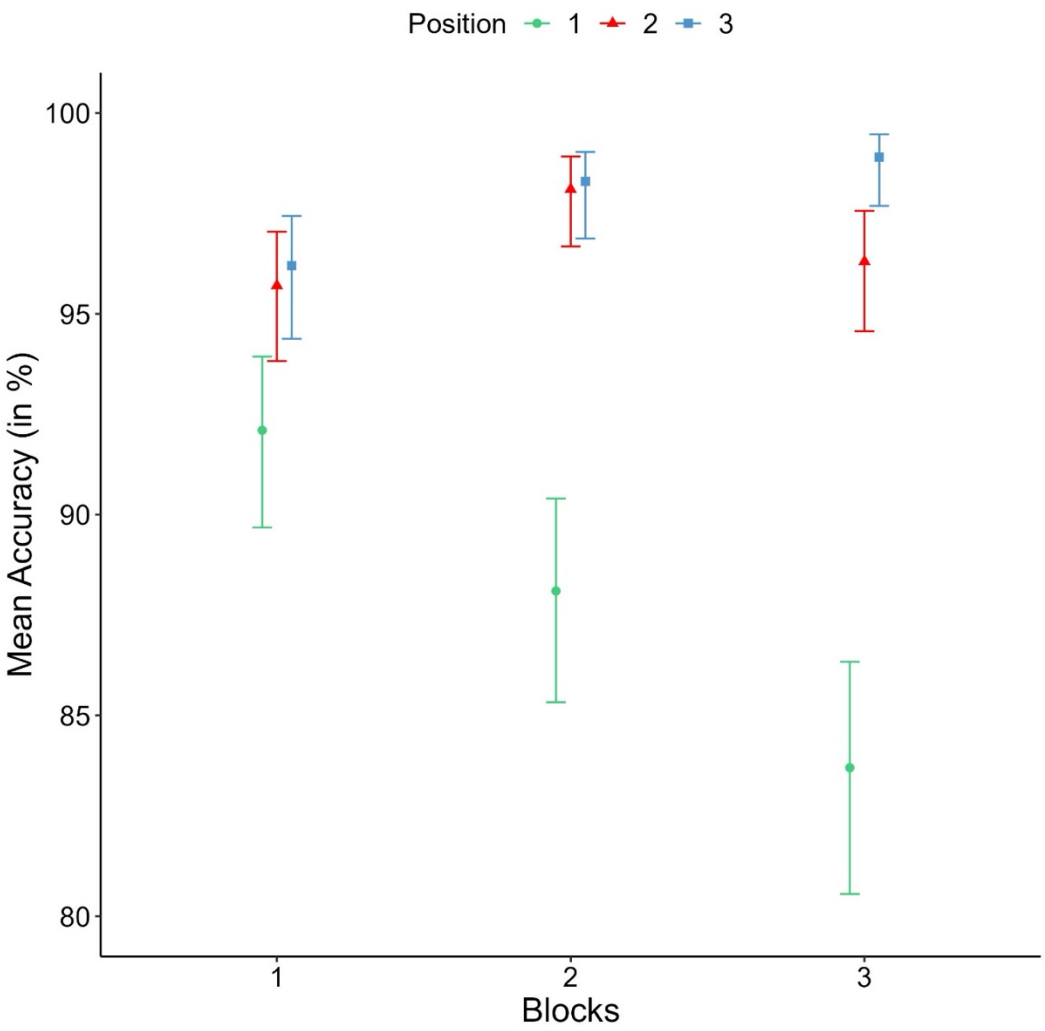


Figure 4. Mean accuracy in Experiment 2 as a function of pseudoword position in the repeated triplet and block number. Error bars indicate 95% confidence intervals.

Discussion

The results of Experiment 2 partly replicated those of Experiment 1. A first main difference between the two experiments concerns the overall slower RTs obtained for pseudowords compared to words: when averaging the RTs of all three positions, the mean RTs on their first occurrence was 654 ms for words and 810 ms for pseudowords; on their last occurrence (i.e., at the 45th repetition), the mean RTs for words was 429 ms and 470 ms for pseudowords. Apart from these longer RTs, the learning dynamics also produced noticeable differences compared with the one observed for words.

Regarding the repetition effect that is measured by the evolution of RTs for pseudowords occurring at Position 1 of the triplet, the dynamics was clearly different compared to words with a fast decrease of response times during the first 5 repetitions (with a mean RT of 798 ms for the first occurrence and of 704 ms for the 5th repetition), followed by a smoother decrease until the last repetition (with a mean RT of 638 ms for the 45th repetition). While the beta coefficient of the first regression line was -4.22 for words, it was much larger for pseudowords (-23.81). The processing gain for pseudowords at Position 1 (i.e., the difference between mean RTs for the last repetition and the first occurrence) was 160 ms, which is much larger than the one obtained for words (90 ms). Pseudowords seem therefore to benefit to a larger extent from the repetition effect indicating that repetitions produced a fast change in the way these pseudowords were processed and in the way their trace developed in memory.

For predictable pseudowords (i.e., in Position 2 and 3 of the triplet), the broken-stick regression analysis also identified two break points that were slightly different from those obtained with words (for pseudowords, 7.44 and 22 at Position 2, and 6.07 and 18.37 at Position 3; for words, 5.35 and 19.81 at Position 2, and 5.88 and 19.54 at Position 3). Apart from these differences, the learning dynamics were similar with a fast decrease in RTs during the initial repetitions followed by an intermediate decrease and a slower one during the last repetitions. Compared to the repetition effect, the predictability effect was again much larger and produced a much stronger processing gain (i.e., for the 3rd Position, when subtracting the mean RTs for the 45th repetition, 390 ms, from the mean RT for the first occurrence, 779 ms, the processing gain was 779-390 = 389 ms).

Contrary to Experiment 1, the data revealed a significant difference between Position 2 and 3, with faster RTs on Position 3 of the triplet. This difference seems to emerge around the same time as in Experiment 1, namely on the sixth repetition of the triplet.

Although this result is consistent with previous finding in sequence learning, here it might be an artifact due to the fact that participants were slower to classify the pseudowords in Position 2 at the beginning of the task, resulting in a higher estimation of the regression intercept compared to the one of Position 3. Due to this unexpected initial difference (that should have been cancelled by the Latin square design), this difference between Position 2 and 3 is difficult to interpret.

Additionally, we found that as the task progressed, it became more difficult for participants to classify the first item of the triplet as being a pseudoword. Indeed, they systematically made more false "yes" judgments for pseudowords in Position 1 than for those in Positions 2 and 3. Interestingly, false "yes" judgments for pseudowords in Position 1 increased over the course of the task, in contrast to those for pseudowords in Positions 2 and 3. This finding, consistent with usage-based theories, suggests that participants gradually became familiar with the first pseudoword of the repeated triplet, which presumably became lexicalised over time. As a result, participants were more likely to respond incorrectly to the first pseudoword in the triplet. Once they recognised the first pseudoword, they simply had to respond correctly to the rest of the triplet. It is worth noting that in Block 3, participants were also more likely to consider the second pseudoword in the triplet to be a word compared to the third, suggesting that the triplet was becoming progressively lexicalised as well.

As for Experiment 1, the number of participants who reported detecting a recurring sequence was high (93%) but the number of participants who were able to fully recall the triplet was much lower (29% in Experiment 2 compared to 38% in Experiment 1). Here again, the data suggest that learning occurred both implicitly and explicitly, and the rate of explicit learning (i.e., with a full recall of the triplet) was lower for pseudowords (29%) than for words (38%).

Overall, Experiment 1 and Experiment 2 yielded similar results regarding the learning dynamics of the repeated triplet, that is, a slower learning rate on the first unpredictable position due to a simple repetition effect, and a much larger learning rate for the predictable positions (i.e., the 2nd and the 3rd) due to the predictability effect. However, in both experiments and contrary to natural language, words and pseudowords were totally unrelated and apart from systematically occurring one after the other, there was no other reason to associate these items. In Experiment 3, we tested whether the use of a triplet composed of semantically related words (e.g., belonging to the same word category, like for example, the fruit category: strawberry, banana, cherry) could have an effect on the learning dynamics of the triplet. We expected semantic relatedness to facilitate learning both at the implicit level (i.e., on RTs) and at the explicit level (i.e., on the recall of the triplet).

Experiment 3

Methods

Participants

Forty-two participants (22 females; M_{age} = 23 years, SD = 4) were paid and recruited via Prolific (www.prolific.co). All participants were native French speaker and reported having no neurological or language disorders. The average LexTALE vocabulary score (Brysbaert, 2013) was 85.08% (SD = 6.37), and no participant was excluded.

Materials

To construct seven semantically related triplets, we selected 21 low-frequency words from the database Lexique 3.83 (New & Pallier, 2020). All words were four-to-six letters monosyllabic or disyllabic singular nouns and had a freqfilms2 frequency ranged from 2 to 10 occurrence per million. Forty-five additional words and 180 pseudowords were selected and used as filler items between two repetitions of the target triplet. All filler

words were monosyllabic or disyllabic singular nouns and were composed of four to six letters. Their freqfilms2 frequency ranged from 10 to 100 occurrences per million. Pseudowords were retrieved from the Lexicon Project (Ferrand et al., 2010), were monosyllabic or disyllabic, and were composed of four to six letters.

A Latin-square design was used, leading to the generation of seven triplets for the 42 participants (i.e., 6 participants per triplet). Every triplet repetition (15 per block) was separated by three to six filler words or filler pseudowords (75 per block). Sixty words and 60 pseudowords were presented in each block. There were an equal number of 'yes' and 'no' responses in the experiment (i.e., 180 for each type of response). Stimuli are listed in Appendix D.

Apparatus and procedure

The apparatus and procedure were identical to that used in Experiments 1 and 2.

Results

Only correct responses were analysed (96.86% of the data). RTs exceeding 1500 ms (1.32% of data) and RTs greater than 2.5 standard deviations from a participant's mean per block and for each of the three possible positions within the triplet (2.26%) were removed. Means and standard deviations per block are shown in Table 6. We constructed a linear mixed-effects model with the maximum random effect structure allowing convergence (Barr, 2013; Barr et al., 2013). This model included Position, Repetition and the interaction term as fixed effects, participant and Item were used as random intercepts with by-participant random slopes for Position. We included word length and log-transformed word frequency for each word in the triplet as covariates. Given that word associations have been shown to influence processing times in multiword sequences (e.g., Carrol and Conklin, 2020), and that the order of presentation of the words in the triplets varied across participants (because of the Latin-squared design), potentially affecting

processing times as some words were more strongly associated than others, we also included a measure of association strength between triplet words as a covariate. As existing free-association databases in French don't contain all the items we used, we decided to calculate the indirect association strength between the words using the JeuxDeMots database (Lafourcade & Joubert, 2008). This database is based on a collaborative online project where participants see a word and provide an association, which is only validated if other peers have suggested the same association. These associations are then weighted according to the number of associations given by the participants to obtain the association strength. To calculate the indirect association strength between two target words, we generated a list of the most frequently associated words with the target word, then selected the most frequent common word between two target words and averaged the association strengths to obtain the indirect association strength measure. For instance, both banana and strawberry were associated with fruit (i.e., 526 and 480, respectively). To obtain the indirect association strength, we then averaged the two values, resulting in an indirect association strength of 503. The results of the model are summarised in Table 7. Figure 5 provides the evolution of mean response times for each position in the triplet and for the 45 repetitions of the triplet.

Table 6

Mean response times (in milliseconds) and standard deviations (in parentheses) for each block in Experiment 3.

	Block 1	Block 2	Block 3
Position 1	613 (115)	581 (110)	576 (103)
Position 2	496 (162)	369 (143)	344 (147)
Position 3	475 (162)	366 (130)	354 (142)

Table 7

Fixed effects of the mixed model for Experiment 3.

Predictors	Estimate	*SE*	95% CI	*p*
(Intercept)	377.62	68.40	[243.57, 511.68]	<.001
Position 2 - 1	-184.91	13.72	[-211.79, -158.03]	<.001
Position 3 - 2	-6.73	7.38	[-21.19, 7.73]	.368
Repetition	-3.56	0.11	[-3.77, -3.35]	<.001
Position 2 - 1 x Repetition	-4.02	0.26	[-4.53, -3.50]	<.001
Position 3 - 2 x Repetition	0.87	0.26	[0.36, 1.38]	<.001
Word length	8.79	10.43	[-11.65, 29.22]	0.411
Word frequency (log)	2.83	10.78	[-18.31, 23.96]	0.798
Association strength bigram 1	0.04	0.08	[-0.11, 0.19]	0.612
Association strength bigram 2	0.08	0.08	[-0.07, 0.23]	0.294

Note. CI: confidence interval; SE: standard error.

The results showed a significant negative effect of Repetition reflecting a decrease in RTs. We also found faster RTs for words in Position 2 compared to those in Position 1, but not to those in Position 3. Finally, there was a significant negative interaction coefficient for the difference between Position 2 and Position 1, and Repetition, and a significant positive interaction coefficient for the difference between Position 3 and Position 2, and Repetition. No significant effects were found for Word length, Word frequency and Association strength for both bigrams. Paired sample *t*-tests comparisons showed that a significant difference between Position 1 compared to Positions 2 and 3 emerged on the third trial, $t(39) = 3.39$, Bonferroni-adjusted $p = .005$.

Following the same procedure as in Experiments 1 and 2, we performed a broken-stick regression on each Position of the repeated triplet. For Position 1, the analysis revealed a breakpoint at repetition 16.72, 95% CI [11.34, 22.10], with RTs decreasing from repetitions 1 to 16.72, $b = -3.96$, 95% CI [-5.80, -2.11], followed by an almost flat slope, $b = 0.01$, 95% CI [-0.76, 0.77]. Concerning Position 2, we estimated two breakpoints at repetitions 4.64, 95% CI [3.08, 6.20] and 20, 95% CI [16.20, 23.80], with RTs rapidly

decreasing from repetitions 1 to 4.64, *b* = -43.01, 95% CI [-63.16, -22.86], continuing to decrease, but at a slower rate from repetitions 4.64 to 20, *b* = -8.85, 95% CI [-11.27, -6.42], followed by a slower decrease until the end of the task, *b* = -1.37, 95% CI [-2.62, -0.11]. For Position 3, we also estimated two breakpoints at repetitions 5.57, 95% CI [3.89, 7.24] and 18.85, 95% CI [14.49, 23.22], with a fast decrease in RTs from repetitions 1 to 5.57, *b* = -35.23, 95% CI [-49.00, -21.47], continuing to decrease at a slower rate from repetitions 5.57 to 18.85, *b* = -7.62, 95% CI [-10.77, -4.46], followed by a slower decreasing until the last repetition, *b* = -0.85, 95% CI [-1.92, 0.21].

Questionnaire

Forty-one of the 42 participants reported noticing a recurrent word sequence; 29 were able to recall the whole triplet, one recalled one subsequence (words 2 and 3), four could recall non-adjacent words (words 1 and 3), five recalled all the words but in the wrong order, and the three remaining could not recall any word.

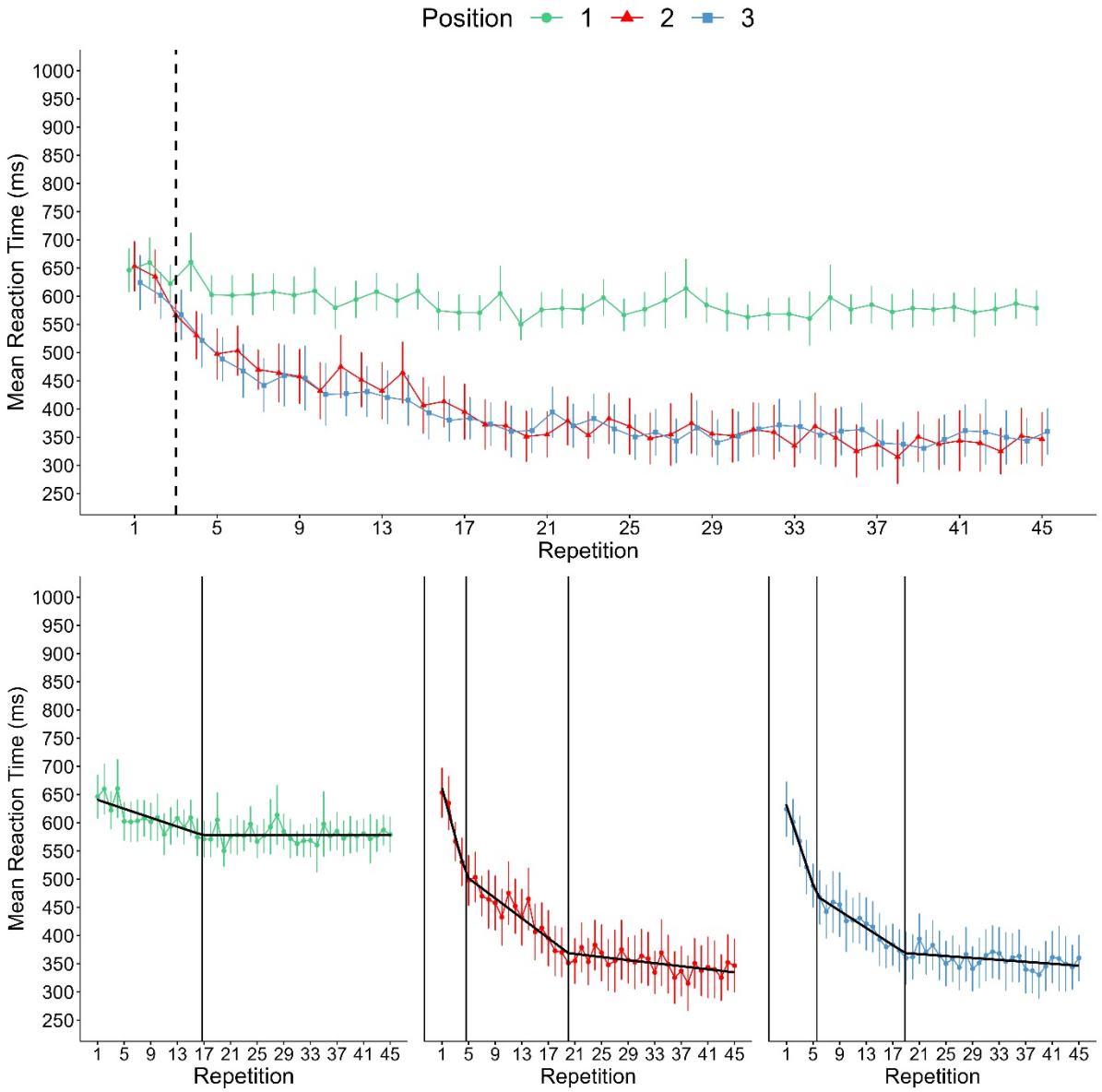


Figure 5. Upper panel: mean response times in Experiment 3 as a function of word position and number of repetitions. The vertical dashed line indicates the first repetition at which there was a significant difference between Position 1 vs. Positions 2 and 3. Error bars indicate 95% confidence intervals. Lower panel: results from the broken-stick regressions for each Position in the triplet. Vertical bars indicate the breakpoints.

Discussion

Experiment 3 produced similar results as in Experiment 1. Concerning the repetition effect, we did not expect any advantage of the semantic relatedness because there is no

reason to observe any effect of this variable on the first word of the triplet. And indeed, the dynamics of the repetition effect was very similar to the one obtained in Experiment 1.

For predictable items (in Position 2 and 3 of the triplet), the beta coefficient of the first regression line (from the broken-stick regression analysis) was larger (-43.01 for related words compared to -32.4 for unrelated words) and the first breakpoint occurred earlier (4.64 compared to 5.35), suggesting that the initial learning phase was much steeper in the semantically related condition compared to the unrelated words from Experiment 1. The semantical relatedness between these words helped producing a larger predictability effect that certainly took advantage of the pre-existing semantic associations between these words. This was also confirmed by the fact that a difference between unpredictable and predictable items emerges earlier than in Experiment 1 (i.e., around the third rather than the fifth repetition of the triplet). Note that this advantage was only present at the early phase of learning because the processing gain for words in Experiment 1 is similar to the one obtained in Experiment 3. Indeed, the difference between the mean response times on Position 3 for the first and last occurrence of these items was 624 ms – 349 ms = 275 ms in Experiment 1 and 624 ms – 360 ms = 264 ms in Experiment 3. Finally, as for Experiment 1, there was no additional advantage for items occurring in Position 3 of the triplet compared to those being in Position 2.

Like Experiment 1, the number of participants who reported detecting a recurring sequence was high (98%) but the number of participants who were able to fully recall the triplet was much larger (69% compared to 38% in Experiment 1). Clearly, the semantic relatedness may have helped participants encoding the triplet in an explicit way which probably also explains the stronger predictability effect observed during the early phase of learning.

As expected, semantic relatedness had a facilitatory effect on the predictability effect but also on the ability of participants to explicitly memorize the repeated triplet and to recall it. However, this situation is rather artificial given that words belonging to the same semantic category rarely appear in a sequence when reading texts, apart from special cases such as binomials (e.g., *salt and pepper*, *boys and girls*, *knife and fork*), which are often composed of words belonging to the same semantic category. It has been shown that the association strength of the component words in binomials influences reading times in a natural reading task (Carrol & Conklin, 2020). We therefore tested whether the learning dynamics of a triplet would be improved by using words that often cooccur, like idioms. A recent study has indeed shown that meaningful three-word sequences (e.g., idioms: on my mind; phrase: is really nice) are easier to process and lead to faster RTs compared to fragment sequences (e.g., because it lets) in a phrasal decision task (Jolsvai et al., 2020). Similarly, Northbrook et al. (2022) presented Japanese English speakers with a series of short stories containing repeated three-word lexical bundles, each seen three times, followed by a phrasal decision task. They found that repeated lexical bundles (e.g., *set off home*, *tired and hungry*) were processed faster than non-repeated bundles in the phrasal decision task, with faster RTs at each subsequent repetition. This advantage for repeated lexical bundles emerged from the first repetition and was still present a week later. In Experiment 4, we therefore used three-word idioms as repeated triplets to study whether the presence of frequently cooccurring words increases the predictability effect. We expected idioms to facilitate learning as they have already been encountered and encoded in memory as whole sequences by the participants.

Experiment 4

Methods

Participants

Forty-two participants (21 females; M_{age} = 24 years, *SD* = 4) were recruited and paid to take part in the study via Prolific (www.prolific.co). All participants were native French speakers and reported having no neurological or language impairments. Their average LexTALE vocabulary score (Brysbaert, 2013) was 86.18% (*SD* = 6.19), no participant was excluded from the analysis.

Materials

We constructed the triplets by selecting seven three-word idiomatic expressions from two databases of French idioms rated by native speakers (Bonin et al., 2013, 2018). Filler items that were inserted between two repetitions of the triplet were 45 words and 180 pseudowords. Words were monosyllabic or disyllabic singular nouns and were chosen from the database Lexique 3.83 (New & Pallier, 2020). All words were four to six letters long and had a freqfilms2 frequency between 10 and 100 occurrences per million. Pseudowords were selected from the French Lexicon Project (Ferrand et al., 2010). All pseudowords were monosyllabic or disyllabic and were composed of four to six letters.

In contrast to previous experiments in which we used a Latin-square design, here the triplets were not scrambled, and therefore participants saw the idioms in their canonical form. Indeed, reversing the word order of existing idiomatic expressions has been shown to result in a processing penalty (Conklin & Carrol, 2020). Each of the 7 idiomatic expressions was presented to six participants (6*7=42). Every triplet repetition (15 per block) was separated by three to six filler words or pseudowords (75 per block). As in the previous experiments, every block was composed of 60 words and 60 pseudowords.

Therefore, there were an equal number of 'yes' and 'no' responses in the experiment (i.e., 180 for each type of response). Stimuli are listed in Appendix E.

Apparatus and procedure

The apparatus and procedure were identical to the one used in Experiments 1, 2 and 3.

Results

Only correct responses were analysed (96.34% of the data). RTs exceeding 1500 ms (1.54% of data), and RTs beyond than 2.5 standard deviations from a participant's mean per block and for each of the three possible positions within the triplet (2.19%) were removed. Mean response times and standard deviations per Block and Position are shown in Table 8.

We constructed a linear mixed-effects model with the maximum random effect structure allowing convergence (Barr, 2013; Barr et al., 2013). This model included Position, Repetition and the interaction term as fixed effects. Item and participant were used as crossed random effects, with by-participant random slopes for Position. In addition to Word length and Word frequency, we also included Idiom frequency and Bigram and Trigram mutual information[2] (MI) scores as covariates in our analysis. Indeed, previous research on idioms has shown that these factors can influence the processing of multiword sequences (e.g., Carrol & Conklin, 2020). Idiom frequency and MI scores were calculated based on the French web corpus frTenTen20 (Jakubíček et al., 2013), which consists of 20.9 billion words. All frequencies were log-transformed prior to analysis. The results of the model are summarised in Table 9. Figure 6 provides the evolution of mean response times for each position in the triplet and for the 45 repetitions of the triplet.

[2] MI estimates the predictability of observing a word given the preceding words in the sequence (Ramisch, 2015, p. 66).

Table 8

Mean response times (in milliseconds) and standard deviations (in parentheses) for each block and position in Experiment 4.

	Block 1	Block 2	Block 3
Position 1	607 (126)	582 (121)	570 (108)
Position 2	478 (179)	337 (156)	309 (136)
Position 3	464 (166)	347 (147)	315 (120)

Table 9

Fixed effects of the mixed model for Experiment 4.

Predictors	Estimate	*SE*	95% CI	*p*
(Intercept)	509.16	84.82	[342.91, 675.42]	<.001
Position 2 - 1	-219.80	33.53	[-285.52, -154.07]	<.001
Position 3 - 2	3.65	12.23	[-20.32, 27.63]	.774
Repetition	-4.09	0.12	[-4.32, -3.87]	<.001
Position 2 - 1 x Repetition	-4.66	0.28	[-5.21, -4.10]	<.001
Position 3 - 2 x Repetition	0.82	0.28	[0.27, 1.38]	.003
Word length	9.43	6.19	[-2.70, 21.56]	.157
Word frequency (log)	-0.01	6.48	[-12.71, 12.69]	.999
Idiom frequency (log)	20.24	9.19	[2.23, 38.25]	.033
Bigram MI	-16.12	4.68	[-25.28, -6.95]	.002
Trigram MI	0.47	4.41	[-8.17, 9.11]	.916

Note. CI: confidence interval; SE: standard error.

The results showed a significant negative effect of Repetition reflecting a decrease in response times with repetitions. We also found faster response times for words in Position 2 compared to those in Position 1, but there was no difference between Position 3 and 2. There were a significant interaction coefficient for the difference between Position 2 and Position 1, and Repetition, as well as for the difference between Position 3 and Position 2, and Repetition. Finally, there was a significant effect of Idiom frequency, with less frequent idioms eliciting faster responses, and of Bigram MI, with faster response times for bigrams with stronger MI. This unusual pattern is most likely due to

the fact that one of our less frequent idioms in the experiment (i.e., *qui dort dîne*) has a high bigram MI score (i.e., 4.72), which may have speeded up participants' responses even though the idiom frequency was low. In fact, any collocation above an MI score of 3 is considered to be strong. When this idiom is excluded from the analysis, the effect of Idiom frequency is no longer significant, b = 16.65, SE = 11.22, 95% CI = [-5.33, 38.64], p = .147. Similarly to Experiment 3, paired sample t-tests comparisons showed a significant difference between Position 1 compared to Positions 2 and 3 on the fourth trial, $t(40)$ = 2.70, Bonferroni-adjusted p = .04.

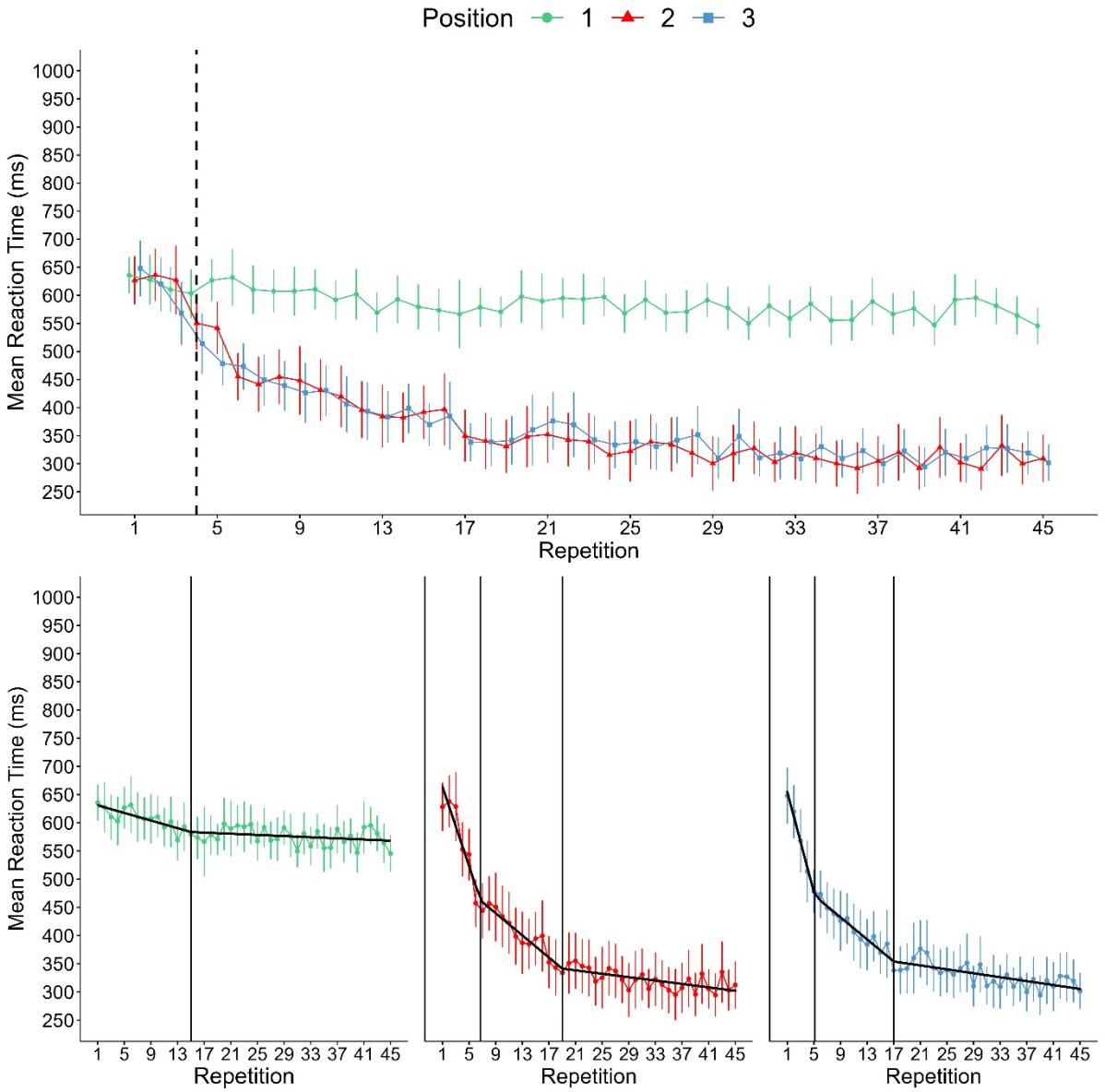


Figure 6. Upper panel: mean response times in Experiment 4 as a function of word position and number of repetitions. The vertical dashed line indicates the first repetition at which there was a significant difference between Position 1 vs. Positions 2 and 3. Error bars indicate 95% confidence intervals. Lower panel: results from the broken-stick regressions for each Position in the triplet. Vertical bars indicate the breakpoints.

We then performed a broken-stick regression to better account for the evolution of the learning pattern throughout the task. A breakpoint was estimated at repetition 15, 95% CI [6.51, 23.49] for Position 1, with RTs decreasing from repetitions 1 to 15, b = -3.38, 95%

CI [-5.81, -0.95], followed by a slower decrease in RTs, *b* = -0.51, 95% CI [-1.25, 0.23]. Regarding Position 2, we estimated two breakpoints at repetitions 6.72, 95% CI [4.88, 8.57] and 19, 95% CI [15.37, 22.63], with a fast decrease in RTs from repetitions 1 to 6.72, *b* = -35.45, 95% CI [-46.72, -24.17], continuing to decrease but at a slower rate from repetitions 6.72 to 19, *b* = -9.87, 95% CI [-13.27, -6.47], followed by a slower decrease until the end of the task, *b* = -1.51, 95% CI [-2.72, -0.31]. For Position 3, two breakpoints were estimated at repetitions 5.17, 95% CI [3.92, 6.41] and 17, 95% CI [13.50, 20.50], with a strong decrease in RTs from repetitions 1 to 5.17, *b* = -44.36, 95% CI [-58.21, -30.51], continuing with a slower decrease from repetitions 5.17 to 17, *b* = -9.79, 95% CI [-13.31, -6.28], followed by an even slower decrease until the end of the task, *b* = -1.74, 95% CI [-2.73, -0.75].

Questionnaire

All participants reported noticing a recurrent word sequence; 37 were able to recall the whole triplet, two recalled one subsequence (words 2 and 3), one could recall non-adjacent words (words 1 and 3), one recalled all the words but in the wrong order, and the last one could not recall any word.

Additional analysis

To compare the predictability effects observed in Experiment 1, 3 and 4, we computed a predictability score for these experiments by calculating a difference between log-transformed RTs for unpredictable words (Position 1) versus the log-transformed mean RT for predictable words (Position 2 and 3) for each repetition. Here, a positive score reflects a predictability effect. We decided to use log-transformed values to control for baseline differences in the participants' responses (see Siegelman, Bogaerts, Kronenfeld, et al., 2018). For instance, let us consider two participants with a mean difference of 100 ms between predictable and unpredictable words, but with a different baseline RT: P1

unpredictable = 600 ms, predictable = 500 ms; P2 unpredictable = 400 ms, predictable = 300 ms. Without this transformation, these participants would have the same difference score, even if the relative acceleration of P2 to predictable words is much higher. After log-transformation, the difference between predictable and unpredictable words reflects better this acceleration: log difference of P1 = 0.18, P2 = 0.29.

We then ran a linear mixed-effects model on the predictability scores, using Experiment, Repetition and the interaction term as fixed effects, and participant as random effect. Experiment was coded using repeated contrast coding (Experiment 1: -0.7 -0.3; Experiment 3: 0.3 -0.3; Experiment 4: 0.3 0.7). We observed higher predictability scores in Experiment 4 (idioms) compared to Experiment 3 (semantically related words), b = 0.07, SE = 0.01, p < .001, and higher scores in Experiment 3 compared to Experiment 1 (non-related words), b = 0.09, SE = 0.01, p < .001. In addition, there was a main effect of Repetition, b = 0.01, SE = 0.00, p < .001, and a significant interaction between Experiment 4 - Experiment 3 and Repetition, b = 0.002, SE = 0.001, p = .008, indicating an increasing difference of predictability scores between both experiments (see Figure 7).

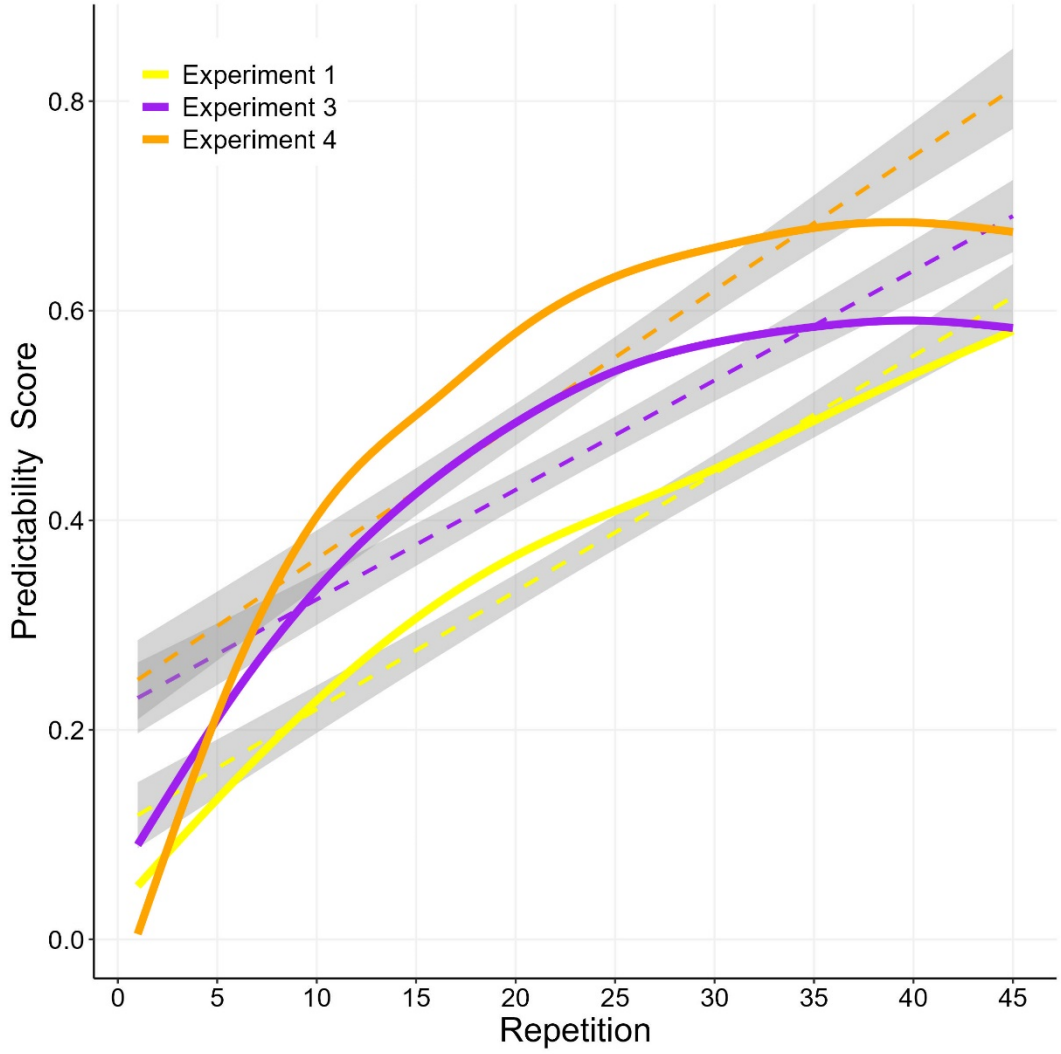


Figure 7. Predictability scores across repetitions for word triplets in Experiments 1, 3 and 4. Continuous lines represent loess fit for the predictability scores. Dashed lines represent the best linear fit and grey-shaded areas indicate 95% confidence intervals around linear regression lines.

Discussion

Experiment 4 produced results similar to Experiment 3. However, two notable differences suggest that idioms have benefited to a larger extent from triplet repetition compared to semantically related words. First, the predictability score represented in Figure 7 indeed shows that when the repetition effect is subtracted from the predictability effect on each repetition trial, the remaining predictability score is stronger for idioms

compared to semantically related words, which is also stronger than the score obtained for unrelated words from Experiment 1. Idioms, which are supposedly already coded in the brain as semantically coherent and frequent sequences of words, appear to derive a greater processing advantage from repetition. Second, while 69% of participants in Experiment 3 were able to recall the full triplet of semantically related words, 88% of participants in Experiment 4 managed to recall the full idiom. This improved performance for explicit correct recall of idioms is probably due to their pre-existing encoding as relevant linguistic sequences, or at least to a facilitated access to them in memory, and it suggests more generally that frequent multiword sequences (apart from idioms) do result in a different learning dynamic in this Hebb lexical decision task compared to less frequent multiword sequences.

General discussion

The goal of the present set of experiments was to provide empirical evidence about the dynamics of multiword sequence extraction by studying the evolution of response times (RTs) for a repeated triplet of items in a task where participants were not informed about the presence of this regularity. Using a Hebb lexical decision task, where a word (Experiments 1, 3 and 4) or a pseudoword (Experiment 2) triplet was repeated throughout a noisy stream of random words and pseudowords, we found that RTs for the unpredictable first position in the triplet decreased over repetitions (i.e., the repetition effect) but more slowly and with a different dynamic compared to items appearing at the predictable second and third positions in the repeated triplet (i.e., the predictability effect). The learning dynamic also varied as a function of triplet type (i.e., unrelated words, pseudowords, semantically related words, or idioms) and there was no evidence of a difference between items appearing at Position 2 and 3 of the triplets. Finally, these results, supported by implicit associative learning mechanisms, were accompanied by

evidence of an explicit learning of the sequence that also varied as a function of the triplet's type.

Repetition is a key mechanism for the development of memory traces for words and sequences of words. There is much recent evidence showing that we acquire not only memory traces for words but also for multiword sequences (e.g., Arnon & Snider, 2010; Bannard & Matthews, 2008; Conklin & Carrol, 2020; Conklin & Schmitt, 2008; Janssen & Barber, 2012; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Siyanova-Chanturia, Conklin, & van Heuven, 2011). The development of these memory traces may facilitate their processing and this phenomenon is now considered by several models of language acquisition (e.g., Abbot-Smith & Tomasello, 2006; Ambridge, 2020; Bannard & Lieven, 2012; McCauley & Christiansen, 2019; Perruchet & Vinter, 1998) as being central for the processing of multiword sequences.

The present set of experiments provides new empirical evidence allowing to better understand the effect of repetitions on the creation of memory traces in the processing of multiword sequences and notably, to differentiate the dynamics of the repetition effect and the predictability effect. The different dynamics of these effects were notably revealed by the broken-stick regression analyses that we conducted on mean response times overall repetitions and for all positions in the repeated triplet. A summary of the main results from these analyses is provided in Table 10.

Table 10

Broken-stick regressions results: breakpoints (BP) and beta coefficients (*b*) of the regression lines for each position in the repeated triplet and for each experiment.

Experiment	Position	BP_1	BP_2	b_1	b_2	b_3
Unrelated words (1)	1	18.46		-4.22	0.52	
	2	5.35	19.81	-32.40	-7.92	-3.20
	3	5.88	19.54	-27.92	-8.18	-1.72
Pseudowords (2)	1	5.18		-23.81	-1.82	
	2	7.44	22.00	-39.33	-10.16	-1.16
	3	6.07	18.37	-41.73	-11.66	-1.86
Semantically related words (3)	1	16.72		-3.96	0.01	
	2	4.64	20.00	-43.01	-8.85	-1.37
	3	5.57	18.85	-35.23	-7.62	-0.85
Idioms (4)	1	15.00		-3.38	-0.52	
	2	6.72	19.00	-35.45	-9.87	-1.51
	3	5.17	17.00	-44.36	-9.79	-1.74

Regarding the repetition effect (indexed by the evolution of RTs on the first unpredictable position of the triplet) for words in Experiment 1, 3, and 4, it was characterized by a late breakpoint (i.e., BP1) and a relatively slow decrease in RTs indexed by a small beta coefficient (i.e., b1). The dynamic was very different for pseudowords in Experiment 2 since it produced an earlier breakpoint (5.18) and a much larger beta coefficient (-23.81). Similarly, the processing gain (indexed by the difference in mean RTs between the 45th repetition and the first occurrence of the item) was smaller for words (i.e., 90 ms, 67 ms, and 90 ms, for Experiment 1, 3, and 4, respectively) than for pseudowords (160 ms). These results suggest that repetition will differentially affect the processing of items that are already encoded in memory (i.e., words) compared to novel items (i.e., pseudowords). Thus, in the present study, we observe that the processing of

novel items benefits very rapidly from repetition and certainly from the transitory development of a memory trace representing these items.

The dynamics of the predictability effect, that is indexed by the evolution of RTs on the second and third predictable positions of the triplet, was characterized, for all items, by a fast decrease in RTs with an early breakpoint (around 4-7 repetitions of the triplet) for the first regression line and a large beta coefficient. The processing gain, which can be computed by subtracting the mean RTs (averaged over Positions 2 and 3) for the last occurrence of the triplet (i.e., 45th repetition) from the mean RTs obtained for the first occurrence of the same items (e.g., 640 ms – 347 ms = 293 ms, for Experiment 1), indicates that the predictability effect was much larger than the repetition effect (i.e., it was 293 ms, 417 ms, 285 ms, and 332 ms, for Experiment 1-4 respectively). The emergence of these early breakpoints for predictable items, as well as of the difference between unpredictable and predictable items (around 3 to 5 repetitions for words), is consistent with the findings of Conklin and Carrol (2020). Indeed, they found a rapid change in participants' reading behaviour after only 4 to 5 repetitions of the repeated pattern. These results clearly illustrate that encoding multiword sequences in memory drastically accelerates the processing of these items and that the predictability effect goes far beyond the repetition effect.

We note that an alternative interpretation to the predictability effect described above can also be provided by the multiconstituent unit (MCU) hypothesis (Zang et al., 2023), which is very close to the assumptions made in McCauley and Christiansen's (2019) computational model. According to this hypothesis, frequently encountered linguistic units consisting of more than a single word can be lexically represented in memory and identified as single representations during reading. In McCauley and Christiansen's (2019) model, this lexicalization process is driven by the central mechanism of chunking

(see also, Perruchet & Vinter, 1998; Jessop et al., 2023). Therefore, multiword and pseudoword sequences that co-occur repeatedly and frequently, as in our study, may gradually become lexicalised and represented as single units in the individual's mental lexicon. Note that several studies by Liang et al. (2015, 2017, 2021, 2023) provide empirical data in favour of this hypothesis in the field of Chinese word reading.

In addition, it is worth noting that the different learning dynamics that we observed for pseudowords can be explained not only by the development of a new memory trace, but also by the lexical decision task itself and the cognitive processes underlying it. Indeed, while in Experiment 2 participants had to give a "no" response to the triplet consisting of pseudowords, it has been shown that producing a "yes" response involves different processes than producing a "no" response. For instance, based on McClelland and Rumelhart's (1981) interactive activation model, Grainger and Jacobs (1996) propose that the generation of a "yes" response occurs when a word is recognised as a result of surpassing a certain activation threshold. In contrast, a "no" response is generated on the basis of global lexical activation, which varies as a function of the likelihood that the stimulus is a word (see also Dufau et al., 2012). Experiment 2 is therefore not comparable to the other experiments in this regard. Nevertheless, like other studies of novel words and multiword sequences using pseudowords (e.g., Norris et al., 2018; Pellicer-Sánchez, 2017; Pellicer-Sánchez et al., 2022; Szmalec et al., 2012), it allows us to study the dynamics of the development of a trace in memory and its influence, in this case, on lexical decision processes. This data may also have direct consequences for computational models of language acquisition like, for example, the Parser model (Perruchet & Vinter, 1998). In this model, each time a unit is processed again (i.e., its processing is repeated), it receives a linear increase of its memory trace (indexed by a weight value). The present results suggest that this increase may not be linear but rather non-linear depending on

the weight of the item's memory trace. For new memory traces, the increase seems to be stronger and more rapid than for memory traces that are more strongly encoded in lexical memory (called "perceptual shaper" in this model).

Likewise, and beyond the repetition effect, the repeated temporal co-occurrence of items provides a strong and non-linear processing advantage for the predictable items. Following Hebbian learning principles (e.g., Brunel & Lavigne, 2009; Endress & Johnson, 2021; Tovar et al., 2018), the coactivation of populations of neurons coding for each item may result in the strengthening of the connection weights between these two populations, leading to the creation of a chunk. Another possibility is to assume that both populations of neurons are activating a third population of pair-coding neurons (Miyashita, 2004) that would code for the pairing of these items. Irrespective of these two possible implementations, the present data suggest that these learning dynamics are non-linear, with a fast development of the memory trace of the chunk followed by a slower regime of memory consolidation.

Although the broken-stick analyses did not permit differentiation of the processing dynamics of the three types of words used in Experiment 1, 3, and 4 (i.e., unrelated words, semantically related words, and idioms, respectively), the predictability scores reported in Figure 6 indicate that the processing of idioms benefited more from the predictability effect than the processing of semantically related words, which also benefited more from the predictability effect than the unrelated words of Experiment 1. This is in line with previous studies showing that prior linguistic knowledge influences and facilitates regularity extraction (e.g., Elazar et al., 2022; Siegelman, Bogaerts, Elazar, et al., 2018). Pre-existing associations between words would then support the predictability effect and notably for idioms which are sequences that are supposedly already represented and supported by memory traces.

It is difficult however to determine whether the advantage for idioms was mainly supported by implicit associative learning or by the participants' prior knowledge of idioms. Table 11 provides a summary of the participants' responses to the final questionnaire, and it clearly suggests that participants' explicit knowledge resulted in stronger learning for idioms compared to semantically related words, which only benefited from implicit learning. Therefore, participants' explicit knowledge of the sequence may have interacted with implicit associative learning mechanisms and the stronger predictability score obtained for idioms may be a product of both factors.

Table 11

Participants' responses to the questionnaire expressed in percentages for each experiment.

Experiment	Participants who noticed a repeated sequence	Participants who correctly recalled the sequence	Participants who did not recall any words of the sequence
Unrelated words (1)	98	38	7
Pseudowords (2)	93	29	31
Semantically related words (3)	98	69	7
Idioms (4)	100	88	2

Finally, the present data did not reveal a processing advantage for the third position over the second, contrary to previous findings on regularity extraction in naming (Rey et al., 2020) and visuomotor tasks (Minier et al., 2016; Rey et al., 2019, 2022). This is likely due to the specific time-scale of the present experimental paradigm that does not allow chunking to occur beyond two items. Indeed, for Hebbian learning to occur between the first and third items in the repeated triplet, it certainly requires maintaining the activation of the neural population coding for the first item long enough to be coactivated with the neural population coding for the last item. However, contrary to previous experimental

paradigms that have reported a learning advantage on the last position of a triplet sequence, lexical decision takes a longer processing time and requires greater attentional load. Both of these factors may lead to a fast deactivation of items that were processed two steps before, avoiding any possible association to occur between item one and three of the repeated triplets. This is consistent with recent findings suggesting that long-distance associations are harder to establish and only occur under very specific conditions (Tosatto et al., 2022; Wilson et al., 2018).

The absence of effect on the third position of the triplets may also be related to a limitation of the present study. Indeed, participants may have learnt two-item associations during the task because stimulus presentation was sequential. It has been argued that parallel presentation is essential for determining the creation of co-word dependencies (Snell et al., 2018). Therefore, sequential presentation might have influenced word extraction and hindered the formation of a 3-word chunk.

In addition, a number of factors are likely to have influenced the learning dynamics during the task, and thus constitute limitations to our study. First, one third of the words forming the triplets in Experiments 1 and 3 can be considered as being part of existing multiword sequences in French (e.g., collocations: "*noyau dur*", "*tronc commun*"; idioms: "*ramener sa fraise*", "*battre le rappel*"). When presented with these words, participants may have already made predictions about the upcoming words in the sequence on the basis of these pre-existing multiword sequences. However, thanks to the counterbalancing on participants, the effect of these pre-existing co-occurrences should be minimized. In addition, most of these existing multiword sequences contain words from parts of speech other than a noun in the first and second positions, making the prediction of the last word in the triplet nearly impossible given the preceding words.

Second, the fact that the triplets in Experiment 4 consisted of different parts of speech (i.e., noun, verb and adjective) compared to those in Experiments 1 and 3 (i.e., nouns) may also have influenced their processing during the task. Indeed, reaction times have been shown to differ across parts of speech (e.g., Kauschke & Stenneken, 2008; Kostić & Katz, 1987; Monaghan et al., 2003; Sereno, 1999; Tyler et al., 2001). Similarly, since the triplets were not matched in terms of MI across the experiments, certain words in some triplets are much less predictive of the following words in the sequence. This is particularly the case in Experiment 4, where the verb *faire* (*to do* in English) is the first word in four triplets. Hence, it may be difficult to directly compare the learning dynamics observed in Experiment 4 with those of the other experiments. Future studies that control for these confounding factors are therefore needed.

Third, given the large number of triplet repetitions (i.e., 45), this task is far from mimicking a real reading situation in which multiword sequences are widely spaced from one another and occur much less frequently. Nevertheless, the use of a well-controlled environment allowed us to characterise the acquisition of multiword sequences in real-time and to investigate in depth the process of word-to-word associative learning in different linguistic settings (i.e., unrelated words, novel words using pseudowords, semantically related words and idioms). To gain a fuller picture of how multiword sequences are acquired, studies employing more ecological presentation conditions, such as those of Conklin and Carrol (2020) and Sonbul et al. (2022), and using different types and larger multiword sequences are needed.

Conclusion

The current study provides novel information about the learning dynamic of multiword sequences when presented in a noisy environment, as is the case in natural language. Our data suggests that multiword learning is carried out through chunking of local information and shows how repetition affects the development of memory traces and improves processing. To further explore and understand the dynamic of multiword sequences extraction, future research could manipulate different parameters from the present experimental Hebb lexical decision task, like for example, the spacing between two repetitions of the repeated sequence or the size of the sequence, to determine the limits of the conditions under which associative learning can occur between a sequence of words.

Ethical approval and Informed Consent

Before starting the experiment, participants accepted an online informed-consent form. Ethics approval was obtained from the "Comité de Protection des Personnes SUD-EST IV" (17/051).

Open Practices Statements

All materials for all experiments, data, and analysis code are available via the Open Science Framework at https://osf.io/h2yns/.

Acknowledgments

We would like to thank two anonymous reviewers for their constructive feedback on previous versions of this paper.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

LPA was supported by a doctoral fellowship of the French Ministry of Higher Education, Research, and Innovation. This research was supported by ERC grant 742141, the Convergence Institute ILCB (ANR-16-CONV-0002), the centre for research in education Ampiric, the CHUNKED ANR project (#ANR-17-CE28-0013-02), the COMPO ANR project (#ANR-23-CE23-0031) and the HEBBIAN ANR project (#ANR-23-CE28-0008).

References

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *Linguistic Review*, *23*(3), 275–290. https://doi.org/10.1515/TLR.2006.011

Ambridge, B. (2020). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, *40*(5–6), 509–559. https://doi.org/10.1177/0142723719869731

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(2), 239–273. https://doi.org/10.1017/S030500091400049X

Angele, B., Baciero, A., Gómez, P., & Perea, M. (2022). Does online masked priming pass the test? The effects of prime exposure duration on masked identity priming. *Behavior Research Methods*, *17*. https://doi.org/10.3758/s13428-021-01742-y

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*(2), 286–304. https://doi.org/10.1111/j.1551-6709.2011.01200.x

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth—Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107–129. https://doi.org/10.1080/15475441.2010.505489

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82. https://doi.org/10.1016/j.jml.2009.09.005

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, *92*, 265–280. https://doi.org/10.1016/j.jml.2016.07.004

Aslin, R. N. (2018). Statistical learning: A powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*, 1–12. https://doi.org/10.1002/wcs.1373.Statistical

Bannard, C., & Lieven, E. (2012). Formulaic Language in L1 Acquisition. *Annual Review of Applied Linguistics*, *32*, 3–16. https://doi.org/10.1017/S0267190512000062

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning of four-word combinations. *Psychological Science*, *19*(3), 241–248. https://doi.org/https://doi.org/10.1111/j.1467-9280.2008.02075.x

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 3–4. https://doi.org/10.3389/fpsyg.2013.00328

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bonin, P., Méot, A., & Bugaiska, A. (2013). Norms and comprehension times for 305 French idiomatic expressions. *Behavior Research Methods*, *45*(4), 1259–1271. https://doi.org/10.3758/s13428-013-0331-4

Bonin, P., Méot, A., Boucheix, J. M., & Bugaiska, A. (2018). Psycholinguistic norms for 320 fixed expressions (idioms and proverbs) in French. *Quarterly Journal of Experimental Psychology*, *71*(5), 1057–1069. https://doi.org/10.1080/17470218.2017.1310269

Borensztajn, G., Zuidema, W., & Bod, R. (2009). Children's grammars grow more abstract

with age-evidence from an automatic procedure for identifying the productive units of language. *Topics in Cognitive Science*, *1*(1), 175–188. https://doi.org/10.1111/j.1756-8765.2008.01009.x

Brunel, N., & Lavigne, F. (2009). Semantic priming in a cortical network model. *Journal of Cognitive Neuroscience*, *21*(12), 2300-2319.

Brysbaert, M. (2013). LEXTALE-FR a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica*, *53*(1), 23–37. https://doi.org/10.5334/pb-53-1-23

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A Tutorial. *Journal of Cognition*, *1*(1), 1–20. https://doi.org/10.5334/joc.10

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711–733.

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, *63*(1), 95–122. https://doi.org/10.1177/0023830918823230

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*. https://doi.org/10.1017/S0140525X1500031X

Conklin, K., & Carrol, G. (2020). Words Go Together Like 'Bread and Butter': The Rapid, Automatic Acquisition of Lexical Patterns. *Applied Linguistics*, *42*(3), 1–23. https://doi.org/10.1093/applin/amaa034

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*,

29(1), 72–89. https://doi.org/10.1093/applin/amm022

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356–371. https://doi.org/10.1016/j.cognition.2009.10.009

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press.

Dufau, S., Grainger, J., & Ziegler, J. C. (2012). How to say "no" to a nonword: A leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1117–1128. https://doi.org/10.1037/a0026948

Elazar, A., Alhama, R. G., Bogaerts, L., Siegelman, N., Baus, C., & Frost, R. (2022). When the "tabula" is anything but "rasa:" What determines performance in the auditory statistical learning task? *Cognitive Science*, *46*(2). https://doi.org/10.1111/cogs.13102

Endress, A. D., & Johnson, S. P. (2021). When forgetting fosters learning: A neural network model for statistical learning. *Cognition*, *213*, 104621.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2010). The French lexicon project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42*(2), 488–496. https://doi.org/10.3758/BRM.42.2.488

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: A unified javascript framework for online studies. *2017 International Conference on Computational Social Science IC2S2*.

Fournet, C., Mirault, J., Perea, M., & Grainger, J. (2022). Effects of letter case on processing sequences of written words. *Journal of Experimental Psychology: Learning, Memory,*

and Cognition, *48*(12), 1–9. https://doi.org/10.1037/xlm0001179

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614–636. https://doi.org/10.1037/a0025255

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, *29*(2), 228–244. https://doi.org/10.1016/0749-596X(90)90074-A

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*(3), 518–565. https://doi.org/10.1037/0033-295X.103.3.518

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning: A Symposium* (pp. 37–46).

Ibbotson, P. (2013). The scope of usage-based theory. *Frontiers in Psychology*, *4*, 1–15. https://doi.org/10.3389/fpsyg.2013.00255

Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2022). Individual differences in artificial and natural language statistical learning. *Cognition*, *225*, 105123. https://doi.org/10.1016/j.cognition.2022.105123

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen corpus family. *7th International Corpus Linguistics Conference CL*, 125–127.

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0033202

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *20*(4), 824–843.

Jessop, A., Pine, J., & Gobet, F. (2023). *Chunk-based Incremental Processing and Learning: An integrated theory of word discovery, vocabulary growth, and speed of lexical processing*. PsyArXiv. https://doi.org/10.31234/osf.io/dukpt

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2020). Meaningfulness beats frequency in multiword chunk processing. *Cognitive Science*, *44*(10), e12885. https://doi.org/10.1111/cogs.12885

Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology*, *98*, 1–21. https://doi.org/10.1016/j.cogpsych.2017.07.002

Kauschke, C., & Stenneken, P. (2008). Differences in noun and verb processing in lexical decision cannot be attributed to word form and morphological complexity alone. *Journal of Psycholinguistic Research*, *37*(6), 443–452. https://doi.org/10.1007/s10936-008-9073-3

Kostić, A., & Katz, L. (1987). Processing differences between nouns, adjectives, and verbs. *Psychological Research*, *49*(4), 229–236. https://doi.org/10.1007/BF00309031

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lafourcade, M., & Joubert, A. (2008). JeuxDeMots: Un prototype ludique pour l'émergence de relations entre termes. *JADT'08 : Journées internationales d'Analyse statistiques des Données Textuelles*, 657–666.

Lenth, R. V. (2023). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.9. https://CRAN.R-project.org/package=emmeans.

Liang, F., Blythe, H. I., Bai, X., Yan, G., Li, X., Zang, C., & Liversedge, S. P. (2017). The role of character positional frequency on Chinese word learning during natural reading. *PLOS ONE*, *12*(11), e0187656. https://doi.org/10.1371/journal.pone.0187656

Liang, F., Blythe, H. I., Zang, C., Bai, X., Yan, G., & Liversedge, S. P. (2015). Positional character frequency and word spacing facilitate the acquisition of novel words during Chinese children's reading. *Journal of Cognitive Psychology*, *27*(5), 594–608. https://doi.org/10.1080/20445911.2014.1000918

Liang, F., Gao, Q., Li, X., Wang, Y., Bai, X., & Liversedge, S. P. (2023). The importance of the positional probability of word final (but not word initial) characters for word segmentation and identification in children and adults' natural Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(1), 98–115. https://doi.org/10.1037/xlm0001116

Liang, F., Ma, J., Bai, X., & Liversedge, S. P. (2021). Initial landing position effects on Chinese word learning in children and adults. *Journal of Memory and Language*, *116*, 104183. https://doi.org/10.1016/j.jml.2020.104183

McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, *126*(1), 1–51. https://doi.org/10.1037/rev0000126

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

Meylan, S. C., Frank, M. C., Roy, B. C., & Levy, R. (2017). The emergence of an abstract grammatical category in children's early speech. *Psychological Science*, *28*(2), 181–192. https://doi.org/10.1177/0956797616677753

Minier, L., Fagot, J., & Rey, A. (2016). The temporal dynamics of regularity extraction in

non-human primates. *Cognitive Science*, *40*(4), 1019–1030. https://doi.org/10.1111/cogs.12279

Mirault, J., Snell, J., & Grainger, J. (2018). You that read wrong again! A transposed-word effect in grammaticality judgments. *Psychological Science*, *29*(12), 1922–1929. https://doi.org/10.1177/0956797618806296

Miyashita, Y. (2004). Cognitive memory: cellular and network machineries and their top-down control. *Science, 306*, 435-440.

Monaghan, P., Chater, N., & Christiansen, M. H. (2003). Inequality between the classes: Phonological and distributional typicality as predictors of lexical processing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*, 810–815.

Mosse, E. K., & Jarrold, C. (2008). Hebb learning, verbal short-term memory, and the acquisition of phonological forms in children. *Quarterly Journal of Experimental Psychology*, *61*(4), 505–514. https://doi.org/10.1080/17470210701680779

Muggeo, V. M. R. (2008). Segmented: An R package to fit regression models with broken-line relationships. *R News*, *8*(1), 20–25. https://cran.r-project.org/doc/Rnews/

New, B., & Pallier, C. (2020). *Lexique 3.83*. http://www.lexique.org

Norris, D., Page, M. P. A., & Hall, J. (2018). Learning nonwords: the Hebb repetition effect as a model of word learning. *Memory*, *26*(6), 852–857. https://doi.org/10.1080/09658211.2017.1416639

Northbrook, J., Allen, D., & Conklin, K. (2022). 'Did you see that?'—The role of repetition and enhancement on lexical bundle processing in English learning materials. *Applied Linguistics*, *43*(3), 453–472. https://doi.org/10.1093/applin/amab063

Ordonez Magro, L., Mirault, J., Grainger, J., & Majerus, S. (2022). Sequential versus simultaneous presentation of memoranda in verbal working memory: (How) does it matter? *Memory and Cognition*. https://doi.org/10.3758/s13421-022-01284-4

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3737–3753. https://doi.org/10.1098/rstb.2009.0173

Page, M. P. A., Cumming, N., Norris, D., McNeil, A. M., & Hitch, G. J. (2013). Repetition-spacing and item-overlap effects in the Hebb repetition task. *Journal of Memory and Language*, *69*(4), 506–526. https://doi.org/10.1016/j.jml.2013.07.001

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, *21*(3), 381–402. https://doi.org/10.1177/1362168815618428

Pellicer-Sánchez, A., Siyanova-Chanturia, A., & Parente, F. (2022). The effect of frequency of exposure on the processing and learning of collocations: A comparison of first and second language readers' eye movements. *Applied Psycholinguistics*, *43*(3), 1–30. https://doi.org/10.1017/S014271642200011X

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685. https://doi.org/10.1111/j.1467-8624.2009.01290.x

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246–263. https://doi.org/10.1006/jmla.1998.2576

Pinker, S. (1991). Rules of language. *Science*, *253*(5019), 530–535. https://doi.org/10.1126/science.1857983

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456–463. https://doi.org/10.1016/S1364-6613(02)01990-3

Ramisch, C. (2015). *Multiword Expressions Acquisition*. Springer. https://doi.org/10.1007/978-3-319-09207-2

Rey, A., Bogaerts, L., Tosatto, L., Bonafos, G., Franco, A., & Favre, B. (2020). Detection of regularities in a random environment. *Quarterly Journal of Experimental Psychology*, *73*(12), 1–13. https://doi.org/10.1177/1747021820941356

Rey, A., Fagot, J., Mathy, F., Lazartigues, L., Tosatto, L., Bonafos, G., Freyermuth, J.-M., & Lavigne, F. (2022). Learning higher-order transitional probabilities in nonhuman primates. *Cognitive Science, 46*(4), e13121. Doi: 10.1111/cogs.13121

Rey, A., Minier, L., Malassis, R., Bogaerts, L., & Fagot, J. (2019). Regularity extraction across species: Associative learning mechanisms shared by human and non-human primates. *Topics in Cognitive Science*, *11*(3), 573–586. https://doi.org/10.1111/tops.12343

Robinet, V., Lemaire, B., & Gordon, M. B. (2011). MDLChunker: A MDL-Based cognitive model of inductive learning. *Cognitive Science*, *35*, 1352–1389. https://doi.org/10.1111/j.1551-6709.2011.01188.x

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 1–17. https://doi.org/10.1037/0096-1523.3.1.1

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*,

110, 104038. https://doi.org/10.1016/j.jml.2019.104038

Schimke, E. A. E., Angwin, A. J., Cheng, B. B. Y., & Copland, D. A. (2021). The effect of sleep on novel word learning in healthy adults: A systematic review and meta-analysis. *Psychonomic Bulletin & Review*, *28*(6), 1811–1838. https://doi.org/10.3758/s13423-021-01980-3

Sereno, J. A. (1999). Hemispheric differences in grammatical class. *Brain and Language*, *70*(1), 13–28. https://doi.org/10.1006/brln.1999.2137

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*(2), 418–432. https://doi.org/10.3758/s13428-016-0719-z

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "learning" in statistical learning: What does an online measure reveal about the assimilation of visual regularities? *Cognitive Science*, *42*, 692–727. https://doi.org/10.1111/cogs.12556

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*(2), 251–272. https://doi.org/10.1177/0267658310382068

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(3), 776–784. https://doi.org/10.1037/a0022531

Skarabela, B., Ota, M., O'Connor, R., & Arnon, I. (2021). 'Clap your hands' or 'take your hands'? One-year-olds distinguish between frequent and infrequent multiword phrases. *Cognition*, *211*, 104612. https://doi.org/10.1016/j.cognition.2021.104612

Smalle, E. H. M., Bogaerts, L., Simonis, M., Duyck, W., Page, M. P. A., Edwards, M. G., & Szmalec, A. (2016). Can chunk size differences explain developmental changes in lexical learning? *Frontiers in Psychology*, *6*, 1–14. https://doi.org/10.3389/fpsyg.2015.01925

Snell, J., & Theeuwes, J. (2020). A story about statistical learning in a story: Regularities impact eye movements during book reading. *Journal of Memory and Language*, *113*, 104127. https://doi.org/10.1016/j.jml.2020.104127

Snell, J., Leipsig, S. Van, & Grainger, J. (2018). OB1-Reader: A Model of Word Recognition and Eye Movements in Text Reading. *Psychological Review*, *125*(6), 969–984.

Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences*, *102*(33), 11629–11634. https://doi.org/10.1073/pnas.0409746102

Sonbul, S., El-Dakhs, D. A. S., Conklin, K., & Carrol, G. (2022). "Bread and butter" or "butter and bread"? Nonnatives' processing of novel lexical patterns in context. *Studies in Second Language Acquisition*, 1–23. https://doi.org/10.1017/S0272263122000237

Szmalec, A., Duyck, W., Vandierendonck, A., Mata, A. B., & Page, M. P. A. (2009). The Hebb repetition effect as a laboratory analogue of novel word learning. *Quarterly Journal of Experimental Psychology*, *62*(3), 435-443. https://doi.org/10.1080/17470210802386375

Szmalec, A., Page, M. P. A., & Duyck, W. (2012). The development of long-term lexical

representations through Hebb repetition learning. *Journal of Memory and Language*, *67*(3), 342–354. https://doi.org/10.1016/j.jml.2012.07.001

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition* (1st ed.). Harvard University Press.

Tosatto, L., Bonafos, G., Melmi, J.-B., & Rey A. (2022). Detecting non-adjacent dependencies is the exception rather than the rule. *PLoS ONE*, *17*(7), e0270580. Doi: 10.1371/journal.pone.0270580

Tovar, Á. E., Westermann, G., & Torres, A. (2018). From altered synaptic plasticity to atypical learning: A computational model of Down syndrome. *Cognition*, *171*, 15-24.

Treiman, R., Gordon, J., Boada, R., Peterson, R. L., & Pennington, B. F. (2014). Statistical learning, letter reversals, and reading. *Scientific Studies of Reading*, *18*(6), 383–394. https://doi.org/10.1080/10888438.2013.873937

Tyler, L. K., Richard, R., Fadili, J., & Moss, H. E. (2001). The neural representation of nouns and verbs: PET studies. *Brain*, *124*(8), 1619–1634. https://doi.org/10.1093/brain/124.8.1619

Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language*, *93*(2), 185–238. https://doi.org/10.1016/j.bandl.2004.10.001

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., vander Kant, A., Smith, K., & Rey, A. (2018). Non-adjacent dependency learning in humans and other animals. *Topics in Cognitive Science*, *12*(3), 843-858.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation

of the current state of play. *Annual Review of Applied Linguistics*, *32*, 231–254. https://doi.org/10.1017/S026719051200013X

Zang, C., Fu, Y., Du, H., Bai, X., Yan, G., & Liversedge, S. P. (2023). Processing multiconstituent units: Preview effects during reading of Chinese words, idioms, and phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001234

Appendix A. Participants' scores and standard deviations on the LexTALE task for each experiment.

Experiment	Vocabulary score	*SD*
Unrelated words (1)	86.53%	5.76
Pseudowords (2)	85.13%	7.08
Semantically related words (3)	85.08%	6.37
Idioms (4)	86.18%	6.19

Appendix B. List of triplets per participant and fillers in Experiment 1

Participant	Triplet
1	armure gilet nectar
2	armure nectar gilet
3	gilet armure nectar
4	gilet nectar armure
5	nectar armure gilet
6	nectar gilet armure
7 … 12	mule proie noeud
13 … 18	virage rasoir festin
19 … 24	livret tirage cloche
25 … 30	graine tronc berger
31 … 36	rappel noyau palace
37 … 42	calcul balai volcan

Filler words

angle armée assaut bague balle base câble canard chute cible coton dent doute fuite grève hasard jardin lion moteur musée neige nuage offre orage ours papier parole pause perte pilote pluie porc prince rideau roue signe site soupe tarte tenue tigre trafic valise violon voisin

Filler pseudowords

abréne acogne acrule acun agarte ainte alire alme altace anet anide antôt appome arine artal arti arut arêle asode atinle augard ausi autoce blose boce borté buge bune carc caruce catail catire cepame cerson cetir chamir charde charon chefet cheler choui chroid chume ciroir claint cluise coble cocère colmel counai crupe cumite cuse damade danfum degite derler ditrer doble drugue dévede elsir engade ensime falber farbe fauf flerse folde forni frone fube fule garsu gitre glac granal gretro grode grort grupe gumble hatour heudit hoire hontôt hougue humice inssir iple ipéth iseau jada japite jutand laitôt latace lavec lieune lipin léble léfile lémece mevail miman molome moléce monner mèlui môvec naille natéme noil nomsi nopore noxe omante oufage ounite pachet paseau pecran pertif piale plaze pleité pordée preper preur preuro psat puif pérée ranu relle renchi renre renume rerile rerise retave ricit rocèle roulip sabe sarie satu sepoce sesin soite sona soral sounir spho spile suine sule supe taivec taute touet touge toutôt trounu tulque tyle uant ucun unate vecote ventin visise vosé voule vrande vêpel âleur ésale êder îcun

Appendix C. List of triplets per participant and fillers in Experiment 2

Participant	Triplet
1	acré mouffe brague
2	acré brague mouffe
3	mouffe acré brague
4	mouffe brague acré
5	brague acré mouffe
6	brague mouffe acré
7 … 12	étrin rompte flais
13 … 18	charpe crêle joine
19 … 24	souffe prompe dige
25 … 30	bloue sprère vauve
31 … 36	fitre plou boge
37 … 42	relet harte loude

Filler words

acier acteur adieu agence aile angle anneau armée assaut auteur bague balle ballon banque barbe base beurre bible bière blague bombe bonté bourse bouton bruit canard carte cesse change chat chaîne choc chute cible cirque client code coffre copain corde coton course crème cuir câble dent destin dette devoir disque douche doute drôle défi départ dîner empire emploi ennui ferme fleuve forme forêt four fuite fumée genou golf grâce grève gâteau génie hasard herbe hiver huile humour image indice jardin joie lame ligne lion loup lycée magie mairie mardi milieu moteur mouche musée nature navire neige neveu nuage objet offre ombre orage ours page pain papier parfum pari parole pause perte pilote plage pluie poche poil pomme porc port poste potion pouce poème presse preuve prince prénom prêtre pêche radio rideau robe rocher rose roue rythme régime salade salon sauce savon scène signe site siège soirée souci soupe sport statue style tabac tante tarte tasse temple tenue texte tigre toile tombe trafic trou troupe trésor tuyau vague valise vallée ventre veste violon vitre voeu voie voisin vote vélo écoute épée

Filler pseudowords

ainte alire altace antôt arut atinle bune charon chefet choui ciroir cocère counai cuse engade ensime falber farbe fauf flerse fube gretro iple iseau jada latace lieune mèlui naille ounite paseau puif renre retave rocèle sepoce sesin soral suine trounu vecote voule vrande âleur ésale

Appendix D. List of triplets per participant and fillers in Experiment 3

Participant	Triplet
1	banane cerise fraise
2	banane fraise cerise
3	cerise banane fraise
4	cerise fraise banane
5	fraise banane cerise
6	fraise cerise banane
7 … 12	coco figue poire
13 … 18	citron tomate oignon
19 … 24	saumon truite requin
25 … 30	pigeon dinde hibou
31 … 36	coyote renard lièvre
37 … 42	cobaye tortue taupe

Filler words

angle armée assaut bague balle base câble canard chute cible coton dent doute fuite grève hasard jardin lion moteur musée neige nuage offre orage ours papier parole pause perte pilote pluie porc prince rideau roue signe site soupe tarte tenue tigre trafic valise violon voisin

Filler pseudowords

abréne acogne acrule acun agarte ainte alire alme altace anet anide antôt appome arine artal arti arut arêle asode atinle augard ausi autoce blose boce borté buge bune carc caruce catail catire cepame cerson cetir chamir charde charon chefet cheler choui chroid chume ciroir claint cluise coble cocère colmel counai crupe cumite cuse damade danfum degite derler ditrer doble drugue dévede elsir engade ensime falber farbe fauf flerse folde forni frone fube fule garsu gitre glac granal gretro grode grort grupe gumble hatour heudit hoire hontôt hougue humice inssir iple ipéth iseau jada japite jutand laitôt latace lavec lieune lipin léble léfile lémece mevail miman molome moléce monner mèlui môvec naille natéme noil nomsi nopore noxe omante oufage ounite pachet paseau pecran pertif piale plaze pleité pordée preper preur preuro psat puif pérée ranu relle renchi renre renume rerile rerise retave ricit rocèle roulip sabe sarie satu sepoce sesin soite sona soral sounir spho spile suine sule supe taivec taute touet touge toutôt trounu tulque tyle uant ucun unate vecote ventin visise vosé voule vrande vêpel âleur ésale êder îcun

Appendix E. List of triplets per participant and fillers in Experiment 4

Participant	Triplet
1 … 6	être mauvais joueur
7 … 12	faire fausse route
13 … 18	qui dort dîne
19 … 24	faire chou blanc
25 … 30	faire profil bas
31 … 36	faire grise mine
37 … 42	montrer patte blanche

Fillers

angle armée assaut bague balle base câble canard chute cible coton dent doute fuite grève hasard jardin lion moteur musée neige nuage offre orage ours papier parole pause perte pilote pluie porc prince rideau roue signe site soupe tarte tenue tigre trafic valise violon voisin

Filler pseudowords

abréne acogne acrule acun agarte ainte alire alme altace anet anide antôt appome arine artal arti arut arêle asode atinle augard ausi autoce blose boce borté buge bune carc caruce catail catire cepame cerson cetir chamir charde charon chefet cheler choui chroid chume ciroir claint cluise coble cocère colmel counai crupe cumite cuse damade danfum degite derler ditrer doble drugue dévede elsir engade ensime falber farbe fauf flerse folde forni frone fube fule garsu gitre glac granal gretro grode grort grupe gumble hatour heudit hoire hontôt hougue humice inssir iple ipéth iseau jada japite jutand laitôt latace lavec lieune lipin léble léfile lémece mevail miman molome moléce monner mèlui môvec naille natéme noil nomsi nopore noxe omante oufage ounite pachet paseau pecran pertif piale plaze pleité pordée preper preur preuro psat puif pérée ranu relle renchi renre renume rerile rerise retave ricit rocèle roulip sabe sarie satu sepoce sesin soite sona soral sounir spho spile suine sule supe taivec taute touet touge toutôt trounu tulque tyle uant ucun unate vecote ventin visise vosé voule vrande vêpel âleur ésale êder îcun

Chapter 3: Language sequence learning and repetition-spacing

Recent research has shown that a new sequence of words can be learned after only a few repetitions, but little is known about how long this sequence can survive in memory without being reinforced. In the present study, we investigated the influence of repetition spacing on the acquisition of word sequences under incidental learning conditions. We used a Hebb lexical decision task in which a triplet of words was systematically repeated every 4th, 7th, 10th, 20th, 30th or 60th trial, depending on the spacing condition. Our results showed that learning was still possible at large repetition spacing (60-apart spacing) and that the forgetting curve appeared to follow a power law. Furthermore, the processing speed of the words was differentially affected by repetition spacing. These results are discussed in relation to recent statistical learning models of language development[3].

[3] This experimental chapter is an adapted version of the submitted manuscript in thesis format. Pinto Arata, L., Ramisch, Carlos., & Rey, A. (Submitted). Language sequence learning and repetition-spacing.

Introduction

When language learners encounter a novel word or sequence of words, they gradually develop a memory trace of this linguistic event. To prevent this memory trace from fading, it needs to be consolidated over time through repetition. Repetition facilitates children's word learning (Mather & Plunkett, 2009; Schwab & Lew-Williams, 2016; Vihman et al., 2014), and learners are more likely to acquire an unfamiliar word when it is frequently repeated (e.g., Elley, 1989; Uchihara et al., 2019; Webb, 2007). Similarly, better performance in comprehension and production is observed for frequent multiword sequences in both children (e.g., Arnon & Clark, 2011; Bannard & Matthews, 2008; McCauley et al., 2021) and adults (e.g., Arnon & Snider, 2010; Carrol & Conklin, 2020; Jolsvai et al., 2020).

From an early age, infants can learn several words very rapidly (Pinker, 1994; Swingley, 2008). For instance, Kay-Raining Bird and Chapman (1998) found that 13- to 16-month-old infants were able to quickly associate novel words with an object after only four repetitions. Similarly, Tamura et al. (2017) exposed 9- to 11-year-old children to written unfamiliar words that were embedded in short stories. They showed that children were able to learn new word forms correctly after just four repetitions, although the learning of their meaning required more repetitions. Recent research further suggests that four to five repetitions of a repeated multiword sequence are sufficient to observe a change in participants' reading behaviour (Conklin & Carrol, 2020; Pinto Arata et al., 2024; Sonbul et al., 2022).

While most research suggests that approximately 4 repetitions are sufficient to learn a novel word or word sequence, less is known about the influence of repetition spacing (i.e., the distance between each repetition of a repeated sequence). Page et al. (2013) conducted one of the few studies that examines the effect of repetition spacing on word

sequence learning. They used a variant of the Hebb repetition task (Hebb, 1961) in which participants were presented with word sequences for immediate serial recall. Unlike the original task, in which each target sequence (the Hebb sequence) is repeated every third trial and interspersed with random filler sequences, the target sequence occurrences were separated by 6, 9, and 12 random fillers. They showed that the learning rate for the repeated target sequence was substantial in all spacing conditions. Furthermore, St-Louis et al. (2019) extended on this study and found that irregular spacing of repetitions had no impact on the learning rate. Thus, these results indicate that learning is still possible when repetitions are relatively spaced, and that memory traces can survive without constant refreshing.

However, in contrast to the Hebb repetition task, language learning is often incidental, and word sequences are not explicitly rehearsed after each repetition. This raises the question of whether the influence of repetition spacing on word sequence learning is the same under incidental learning conditions. To address this question, we used a Hebb lexical decision task in which participants had to read letter strings presented one at a time on a computer screen and classify them as either words or pseudowords. Unknown to them, a triplet of words (hereafter referred to as ABC) appeared systematically in the same order. Between two repetitions of the ABC triplet, a variable number of random filler words and pseudowords were presented, depending on the spacing condition. We inserted either 4±1 filler items (4F-Condition), 7±1 filler items (7F-Condition), 10±1 filler items (10F-Condition), 20±1 filler items (20F-Condition), 30±1 filler items (30F-Condition), or 60±1 filler items (60F-Condition). Response times (RTs) for each word in the triplet and their evolution over repetitions provide behavioural indicators of the efficiency of their processing and memorisation.

More precisely, the evolution of response times for the first item of the triplet (i.e., A) can be used to study the mere *repetition* effect, as this word is unpredictable due to the random number of filler items occurring between two repetitions of the triplet (Pinto Arata et al., 2024). Similarly, the evolution of response times on the second and third positions of the triplet (i.e., B and C) makes it possible to study the *predictability* effect, i.e. the even greater reduction in response times on predictable positions after several repetitions of the ABC triplet (Rey et al., 2020). The goal of the present study is to investigate how the repetition and predictability effects vary as a function of repetition spacing, and whether these effects are still observed when repetition spacing is relatively large. To our knowledge, this is the first study to test such a large repetition spacing (i.e., up to 60 items).

Experiment

Methods

Participants

Five hundred and thirty-five participants were initially recruited via Prolific (https://www.prolific.com). Out of the 535 participants, we excluded 31: one due to experimental errors, and 30 due to low performance in the language proficiency test described below. Hence, the final sample consisted of 504 participants (279 females; M_{age} = 26.6 years, SD_{age} = 5.6) that were randomly assigned to one of the six spacing conditions (4F, 7F, 10F, 20F, 30F, and 60F), leading to 84 participants per spacing condition. All participants reported being native French speakers with no history of neurological or language impairment. Before starting the experiment, participants signed an online informed consent form.

As participants were recruited online, their proficiency in French was checked using the LexTALE language proficiency test before starting the main task (Brysbaert, 2013).

This test consists of a lexical decision task in which participants are presented with 84 single-item trials (56 real French words, 28 French-looking pseudowords), and are instructed to decide whether each presented letter string is a real French word or not. Any participant whose score was below 2.5 standard deviations from the average LexTALE vocabulary score was excluded from the experiment and replaced. Based on this criterion, thirty participants were excluded and replaced. A summary of the participants' scores and standard deviations on the LexTALE task for each condition is provided in Appendix A.

Materials

A set of 813 words and 882 pseudowords were used as items in this experiment. All words were monosyllabic or disyllabic nouns. They were composed of four to six letters and were selected from the French database Lexique 3.83 (New & Pallier, 2020). Each word of the triplet had a freqfilms2 frequency ranging from 2 to 10 occurrences per million. Filler words had a frequency ranging from 10 to 100 occurrences per million. Pseudowords were obtained from the French Lexicon Project (Ferrand et al., 2010). They were monosyllabic or disyllabic and had a length ranging from four to six letters.

To prevent any order effect, a Latin-square design was used so that each word appeared in every possible position within the triplet, resulting in six possible permutations of each word triplet (ABC, ACB, BAC, BCA, CAB, CBA). Seven triplets were used (for a total of 7*3 = 21 different words), and each of these triplets generated six possible permutations, yielding a total of 7*6 = 42 ordered three-word sequences. Each participant saw a single repeated triplet, with the three words always in the same order. To ensure that participants were presented with an equal number of words and pseudowords, they completed 400 trials in the 4F, 7F and 10F conditions, with 57, 40 and 30 repetitions of the triplet, respectively, whereas they completed 690, 990 and 1764

trials in the 20F, 30F and 60F conditions, with 30 repetitions of the triplet. Due to the large number of trials in the 60F-Condition, this condition was divided into two experimental blocks with a break in between.

To obtain the recommended 1,600 measurements per level of a variable for sufficient statistical power (Brysbaert & Stevens, 2018), we collected the double of the data for each level, that is, each triplet was seen by two participants, resulting in a total of 84 participants per spacing condition. The set of word triplets and fillers is listed in Appendix B.

At the end of the task, all participants answered a short questionnaire. The first question was: “Did you notice anything particular in this experiment?” If the answer was "yes", the follow-up question was “Can you explain what you noticed?”. They were then asked “Did you notice that a sequence of words was systematically repeated?” and “Can you recall the words in their correct serial order?”. If the response to the first question was “No”, participants were only presented with the last two questions.

Apparatus

The experiment was created using LabVanced (Finger et al., 2017) and participants were recruited via the Prolific platform. Participants participated via their personal computer, with only PC-Windows, PC-Linux, and MacOS operating systems allowed, to keep the experimental conditions as similar as possible across participants. All words and pseudowords were presented in lowercase in the centre of the screen using a 20-point Lato black font on a white background.

Procedure

Before starting the experiment, instructions were displayed on the screen. Participants were instructed to decide as fast as possible whether the sequence of letters displayed on the screen formed a French word or not. They had to press “M” (for words) or “Q” (for

pseudowords) on their keyboards, located at the left and right ends of French AZERTY keyboards[4]. RTs and accuracy were recorded for each word and pseudoword. Each target remained on the screen until the participant responded. The next target then appeared immediately after the participant's response. Each spacing condition, as well as each block in the 60F-Condition, began and ended with three filler items to ensure that participants were presented with the triplet at the same time. 4F, 7F and 10F conditions lasted approximately 10 minutes, while 20F, 30F and 60F lasted approximately 12, 16 and 30 minutes, respectively. Figure 1 provides a schematic description of the experimental paradigm.

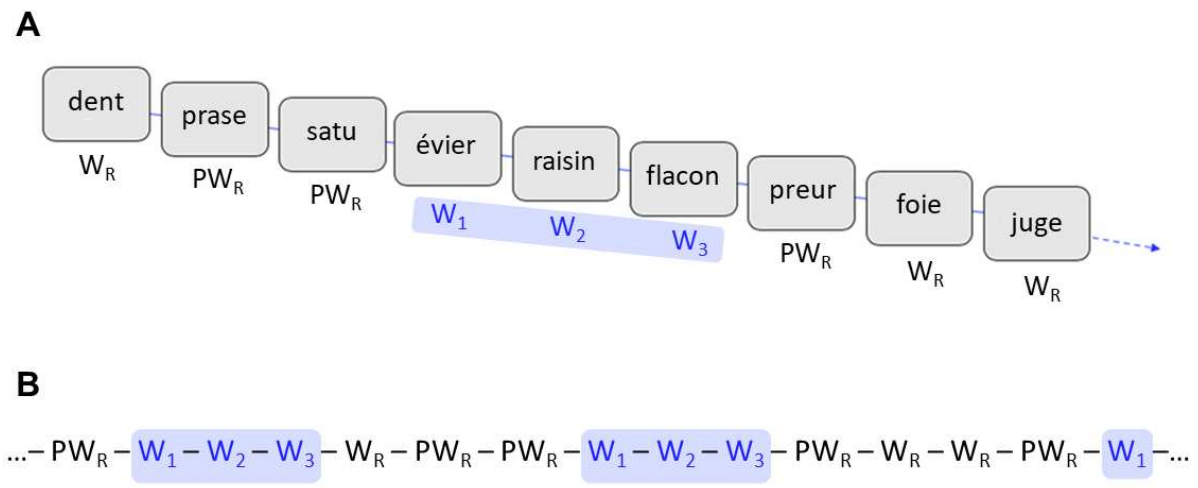


Figure 1. Experimental procedure for the lexical decision task. Panel A: items were presented one at a time in the centre of the screen. Participants had to classify each string as a word or a pseudoword. A repeated sequence of three words (e.g., W_1: "évier" – sink ; W_2: "raisin" – grape ; W_3: "flacon" – flask), always appearing in the same order, was interspersed with random filler words (W_R) or random filler pseudowords (PW_R). Words in blue belong to the repeated triplet. Panel B: one triplet of words ($W_1W_2W_3$) is repeated

[4] On QWERTY keyboards, this would be "A" for words and "L" for pseudowords.

several times and a variable number of random words or pseudowords (W_R or PW_R) are presented between two repetitions of the triplet. The number of random items varied depending on the spacing conditions.

Results

Participants' mean accuracy level was 98.68% in the 4F-Condition, 98.92% in the 7F-Condition, 99.13% in the 10F-Condition, 99.22% in the 20F-Condition, 99.02% in the 30F-Condition, and 98.94% in the 60F-Condition. Incorrect trials were removed from the analysis. We excluded RTs exceeding 1500 ms (0.42%, 0.98%, 0.71%, 0.58%, 0.71%, and 1.19% of the data for conditions 4F, 7F, 10F, 20F, 30F and 60F). To make the conditions comparable, only the first 30 repetitions of the triplet for each condition were included in the analysis. Thus, the final data set consisted of 30 repetitions * 84 participants = 2,520 data points per level. The mean RTs and standard deviations for each position in each condition are presented in Table 1.

To examine the learning dynamics of the repeated multiword sequences, we first compared the evolution of RTs at each of the three positions in the sequence. Data analysis was performed with the R software (version 4.3.2) using linear mixed-effects models (LMEs) fitted with the lme4 package (version 1.1-35.1; Bates et al., 2015) and the lmerTest (version 3.1-3; Kuznetsova et al., 2017). Separate models were fitted for each spacing condition. Each model included the maximum random structure that allowed convergence (Barr, 2013; Barr et al., 2013), that is, Position (1 to 3), Repetition (1 to 30) and their two-way interaction as fixed effects, participants and item sets were used as random effects[5]. Position was coded using Helmert contrasts (i.e., P1: 0.7 0.0; P2: -0.3 0.5; P3: -0.3 -0.5) to compare Position 1 with both Positions 2 and 3, simultaneously, and

[5] The model structure, as denoted in R syntax, was as follows for each model: log(RT) ~ Position*Repetition + (Repetition + Position|Subject) + (1|Item).

Position 2 with Position 3. Repetition was mean centred to reduce collinearity. RTs were log-transformed to reduce skewing and to control for baseline differences between the participants' responses (see Siegelman et al., 2018). The results of the comparison between the positions of the triplet are shown in Table 2.

Table 1

Mean response times in milliseconds (rounded) and standard deviations (in parentheses, rounded) for each position in the triplet and in each spacing condition.

	4F	7F	10F	20F	30F	60F
Position 1	628 (143)	635 (153)	598 (145)	607 (147)	601 (137)	606 (151)
Position 2	468 (183)	534 (195)	491 (173)	506 (167)	533 (152)	531 (167)
Position 3	473 (173)	512 (182)	471 (174)	485 (181)	499 (167)	491 (171)

Table 2

Fixed effect values of each mixed model for all spacing conditions.

	4F			7F			10F		
Predictors	β	SE	*p*	β	SE	*p*	β	SE	*p*
Intercept	6.20	0.02	<.001	6.27	0.02	<.001	6.20	0.02	<.001
P1 vs P2-P3	0.33	0.02	<.001	0.23	0.02	<.001	0.26	0.02	<.001
P2 vs P3	-0.02	0.01	.082	0.05	0.01	<.001	0.05	0.01	<.001
Repetition	-0.01	0.00	<.001	-0.01	0.00	<.001	-0.01	0.00	<.001
P1 vs P2-P3:Repetition	0.01	0.00	<.001	0.01	0.00	<.001	0.01	0.00	<.001
P2 vs P3:Repetition	0.00	0.00	<.001	0.00	0.00	.995	0.00	0.00	.747

Table 2 (continued)

	20F			30F			60F		
Predictors	β	SE	*p*	β	SE	*p*	β	SE	*p*
Intercept	6.22	0.02	<.001	6.26	0.01	<.001	6.25	0.01	<.001
P1 vs P2-P3	0.24	0.02	<.001	0.18	0.02	<.001	0.20	0.02	<.001
P2 vs P3	0.05	0.01	<.001	0.08	0.01	<.001	0.09	0.01	<.001
Repetition	-0.01	0.00	<.001	-0.01	0.00	<.001	-0.01	0.00	<.001
P1 vs P2-P3:Repetition	0.01	0.00	<.001	0.01	0.00	<.001	0.01	0.00	<.001
P2 vs P3:Repetition	0.00	0.00	.031	0.00	0.00	.609	0.00	0.00	.005

Note. P: position, SE: standard error.

Results for all spacing conditions are presented in Figure 2. We found a significant effect of Repetition with an overall decrease of RTs across all spacing conditions. As expected, RTs for predictable words (i.e., on Positions 2 and 3) were significantly faster than those for unpredictable words (i.e., Position 1) in all spacing conditions. In addition, there was a significant positive interaction coefficient for the difference between Position 1 and the average of Positions 2 and 3, and Repetition, indicating that the difference between predictable and unpredictable words increased across repetitions. We also found faster RTs for words in Position 3 compared to those in Position 2 in all the spacing conditions, except for the 4F condition where there was no difference (see Figure 3).

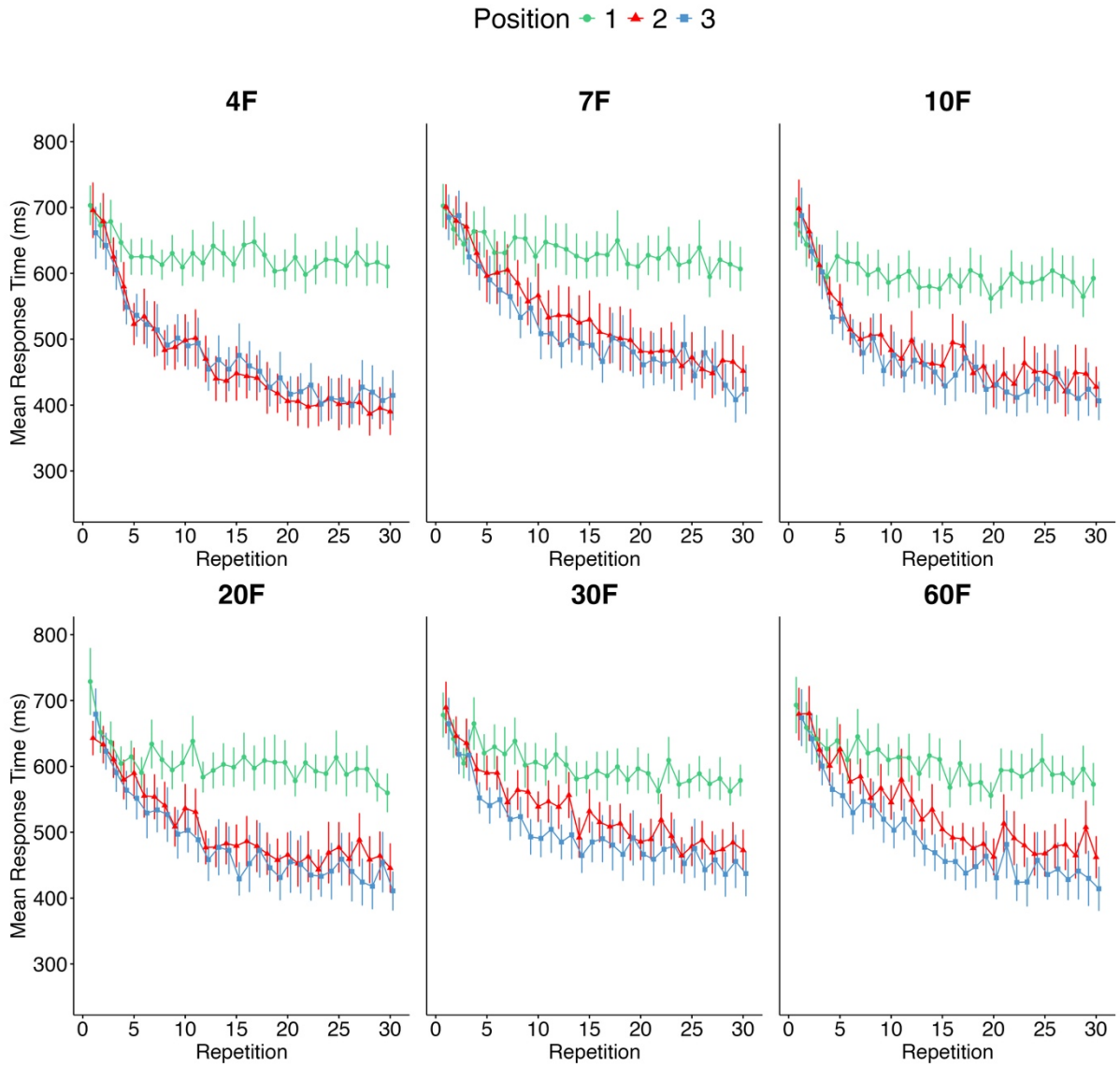


Figure 2. Mean response times in all six spacing conditions as a function of word position and number of repetitions. Error bars indicate 95% confidence intervals.

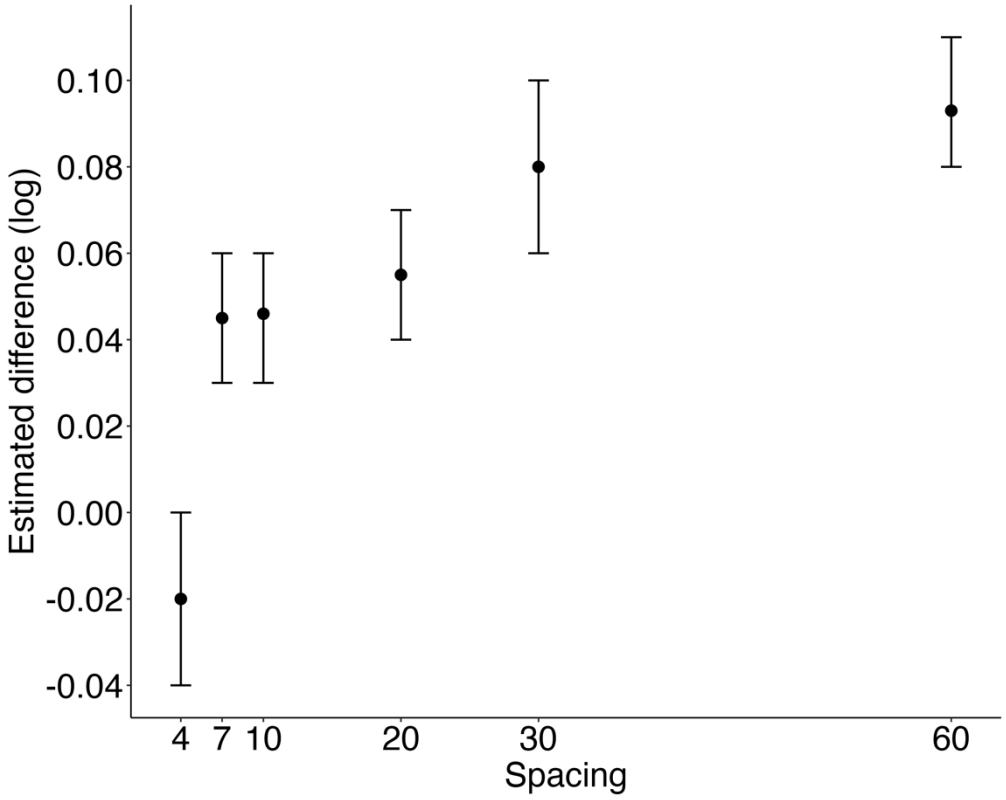


Figure 3. Estimated difference in response times (in log) for words in Positions 2 and 3 from the mixed-effects model, for all six spacing conditions. Error bars represent 95% confidence intervals.

To examine the influence of repetition spacing on the repetition effect more precisely, we ran a model on the RTs for Position 1 with Condition, Repetition, and the interaction term as fixed effects, and participants as random effect. Condition was coded using repeated contrast coding to perform pairwise comparisons between spacing conditions (Schad et al., 2020). We found no difference between the 4F- and 7F-Conditions, slower RTs for the 7F-Condition compared to the 10F-Condition (b = -0.06, SE = 0.02, p = <.001), and no differences between the 10F- and subsequent spacing conditions. This suggests that the repetition effect increases with larger spacing but remains stable after 10-apart spacing.

Next, to examine the effect of repetition spacing on the predictability effect, we computed a predictability score for each spacing condition by calculating the difference

between the log-transformed RTs for unpredictable words (Position 1 of the triplet) versus the log-transformed mean RT for predictable words (Positions 2 and 3 of the triplet). Therefore, a positive score reflects learning of the predictable positions. We then ran a linear mixed-effects model on the predictability scores using the same fixed and random effects as for the repetition effect. The results showed higher predictability scores in the 4F-Condition compared to 7F-Condition, b = -0.10, SE = 0.03, p <.001, and higher scores in the 20F-Condition compared to the 30F-Condition, b = -0.06, SE = 0.03, p = .04. There was no difference in terms of predictability between the 7F-,10F- and 20F-Condition, nor between the 30F- and 60F-Condition.

The predictability effect is particularly important since it informs us about the process of word sequence learning. To get a broader picture of the evolution of the predictability effect over repetition spacing, we extracted the beta coefficients of each linear regression calculated in the mixed model of the predictability effect, and we performed an analysis to determine which regression model best fits the trend observed for the predictability effect. The results showed that the evolution of the predictability effect was better captured by a power law than by a linear, logistic, or exponential regression, R^2 = 0.85, $F(1, 4) = 9.68$, p = .036. It should be noted, however, that this power-law shape is based on a limited number of points, and that the 60F-Condition significantly influenced the regression fit. Figure 4 shows the results of the linear mixed model for the predictability score across the six spacing conditions along with the power regression for the distribution of the beta coefficients. A summary of the other regression models that were fitted is provided in Appendix C.

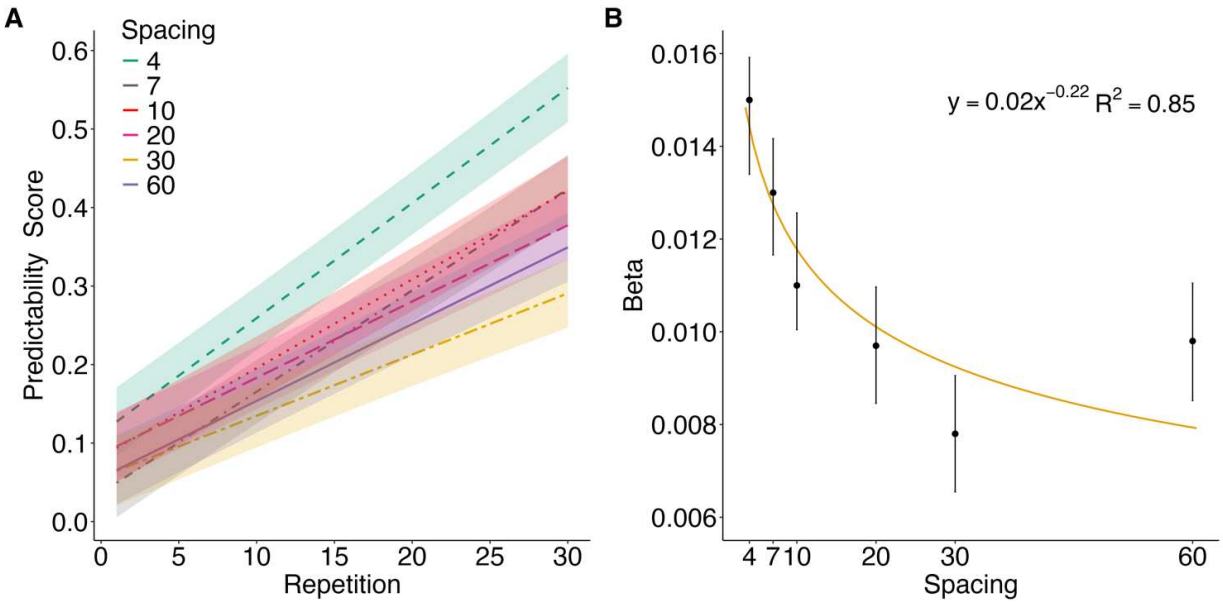


Figure 4. Panel A: Comparison of predictability scores across repetitions for word triplets in all the spacing conditions. Continuous lines represent the best linear fit and colour-shaded areas indicate 95% confidence intervals around linear regression lines. Panel B: Beta coefficients for all six spacing conditions. The continuous line represents the best fit to the data (i.e., a power law). Error bars indicate 95% confidence intervals for each beta coefficient based on the linear mixed model used to compare spacing conditions.

Questionnaire

Among the 84 participants, 48 accurately remembered the entire triplet in the 4F-Condition, 28 in the 7F-Condition, 35 in the 10F-Condition, 32 in the 20F-Condition, 16 in the 30F-Condition, and 20 in the 60F-Condition.

General discussion

In the present study, our primary aim was to investigate the influence of repetition spacing on the incidental learning of multiword sequences. More specifically, we focused on its influence on the repetition and predictability effects, both of which are related to the development of novel memory traces. To this end, we used a Hebb lexical decision task in which a repeated ABC word-triplet was embedded in a random noisy stream.

Unknown to the participants, the ABC word triplet appeared systematically every 4, 7, 10, 20, 30, or 60 trials, depending on the spacing condition. First, we found that the repetition effect was influenced by repetition spacing, with RTs for the unpredictable first position in the triplet decreasing faster across repetitions for spacings longer than 10 items. Second, RTs for the predictable second and third positions in the triplet also decreased across repetitions, and the magnitude of the predictability effect decreased as repetition spacing increased. Finally, based on our regression analysis, we found that the evolution of the predictability effect appeared to best fit a power law.

Regarding the repetition effect, which is associated with enhanced processing due to mere exposure to a word, our findings are consistent with related studies showing that frequently encountered words are processed more efficiently (Brysbaert et al., 2018). Indeed, participants responded faster to the unpredictable first word in the triplet as the number of repetitions increased. Interestingly, this processing advantage was greater for spacings wider than 10 items, which was unexpected. One possible explanation for this finding is that participants' expectations about the occurrence of the repeated sequence may have influenced their performance. In fact, it has been shown that participants who tend to increasingly expect the occurrence of a given stimulus produce faster responses (Destrebecqz et al., 2019). Therefore, as soon as participants noticed that a multiword sequence was repeated, they may have started to anticipate the next occurrence of the sequence, and it is likely that this anticipatory effect had more time to build up in the more spaced conditions, as the time elapsed between sequence repetitions was longer.

Concerning the predictability effect, our results are in line with previous findings on regularity extraction, which have consistently shown a processing advantage for predictable items embedded in a repeated sequence (e.g., Pinto Arata et al., 2024; Rey et al., 2020; Wang et al., 2023). Indeed, participants were faster overall to respond to the

predictable second and third words than to the unpredictable first word in the triplet across spacing conditions. Crucially, our results showed that this processing advantage was modulated by repetition spacing, specifically decreasing when repetition spacing increased. It is worth noting, however, that learning was still substantial even in the 60-Condition. This novel result shows that, with the largest spacing we were able to test, participants are still able to encode the word sequence and extract the statistical relationships between the words in the ABC triplet.

Furthermore, we found a greater processing advantage for the third word over the second, suggesting that the last word in the triplet benefited from the richer contextual information provided by the two preceding words. Surprisingly, this processing advantage for the third word emerged in the 7F-Condition and increased with repetition spacing. One potential explanation for this intriguing phenomenon could come from the evolution of the repetition effect (measured from the evolution of response times on the first position). Since processing times for the first word of the triplet tend to be faster as spacing increases, potentially due to higher expectations, the influence of the first word A on the predictability of words B and C increases. Thus, B's processing time benefits from the improvement in A's processing time, and C's processing time would benefit from the cumulative improvement in A and B's processing times. Further research is needed to better understand this unexpected phenomenon.

Our findings also suggest that the evolution of the predictability effect follows a power law with a small initial forgetting rate, which increases as spacing increases. Interestingly, while the effect of time on retention has been shown extensively (e.g., Averell & Heathcote, 2011; Murre & Dros, 2015; Nickl & Bäuml, 2023), forgetting has been largely overlooked in chunking models of language acquisition. Several recent chunk-based models even assume that there is no forgetting during learning (e.g., Cabiddu et al., 2023; Jessop et al.,

2023; McCauley & Christiansen, 2019; Monaghan & Christiansen, 2010), resulting in each item being permanently stored in the model. One of the few models that includes a forgetting parameter is PARSER (Perruchet & Vinter, 1998). In this model, each time a unit is processed, it receives a fixed increase of its memory trace (represented by a weight value). Conversely, each time a unit remains unprocessed, its memory trace decreases according to a linear trend, leading to complete forgetting of an item after 20 processing steps (corresponding to a 20-apart spacing). Consequently, PARSER simulations are in contradiction with our results, since we observed substantial learning even in the 60F-Condition.

It is also unclear how chunk-based models could account for the repetition and predictability effects. For instance, PARSER (Perruchet & Vinter, 1998) and the CBL model (McCauley & Christiansen, 2019) can extract chunks of different sizes over time, yet there is nothing in their parameters to indicate whether differential processing should be assumed for each word embedded in a chunk. On the other hand, CIPAL (Jessop et al., 2023) implements right-to-left learning, leading to a processing advantage for the last word in the chunk, thereby replicating the predictability effect. Finally, TRACX (French et al., 2011) assumes a left-to-right learning of the regularities which is inconsistent with the present results showing faster learning for words in Position 3 of the triplet relative to words in Position 2 (see Minier et al., 2016).

In summary, our study showed that the incidental learning of multiword sequences is still possible at large repetition spacing. Furthermore, we found that the repetition and predictability effects were influenced by repetition spacing, with a larger repetition effect and a smaller predictability effect at larger spacings. More importantly, our results suggest that, contrary to recent computational models, forgetting plays a critical role in sequence learning and is best fitted by a power law. Finally, participants' performance on

the different positions of a repeated sequence of words seems to be affected differently as the spacing between repetitions increases. Taken together, these findings provide new empirical constraints for statistical learning models of language acquisition.

Ethical approval and Informed Consent

Before starting the experiment, participants accepted an online informed-consent form. Ethics approval was obtained from the "Comité de Protection des Personnes SUD-EST IV" (17/051).

Open Practices Statements

All materials for all experiments, data, and analysis code are available via the Open Science Framework at https://osf.io/guc5d/?view_only=b9c7ec54afed460d8157308135c6c9a9

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research was supported by the Convergence Institute ILCB (ANR-16-CONV-0002), the Center for Research in Education Ampiric, the COMPO ANR project (#ANR-23-CE23-0031), the HEBBIAN ANR project (#ANR-23-CE28-0008), and the SELEXINI ANR project (#ANR-21-CE23-0033-01).

References

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth—Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107–129. https://doi.org/10.1080/15475441.2010.505489

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82. https://doi.org/10.1016/j.jml.2009.09.005

Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology*, *55*(1), 25–35. https://doi.org/10.1016/j.jmp.2010.08.009

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning of four-word combinations. *Psychological Science*, *19*(3), 241–248. https://doi.org/10.1111/j.1467-9280.2008.02075.x

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 3–4. https://doi.org/10.3389/fpsyg.2013.00328

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Brysbaert, M. (2013). LEXTALE-FR a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica*, *53*(1), 23–37. https://doi.org/10.5334/pb-53-1-23

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 1–20. https://doi.org/10.5334/joc.10

Cabiddu, F., Bott, L., Jones, G., & Gambi, C. (2023). Classic utterance boundary: A chunking-based model of early naturalistic word segmentation. *Language Learning*, 1–34. https://doi.org/10.1111/lang.12559

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, *63*(1), 95–122. https://doi.org/10.1177/0023830918823230

Conklin, K., & Carrol, G. (2020). Words go together like 'bread and butter': The rapid, automatic acquisition of lexical patterns. *Applied Linguistics*, *42*(3), 1–23. https://doi.org/10.1093/applin/amaa034

Destrebecqz, A., Vande Velde, M., San Anton, E., Cleeremans, A., & Bertels, J. (2019). Saving the Perruchet effect: A role for the strength of the association in associative learning. *Quarterly Journal of Experimental Psychology*, *72*(6), 1379–1386. https://doi.org/10.1177/1747021818791079

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, *24*(2), 174–184. https://doi.org/10.2307/747863

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2010). The French lexicon project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42*(2), 488–496. https://doi.org/10.3758/BRM.42.2.488

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). Labvanced: A unified Javascript framework for online studies. *2017 International Conference on Computational Social Science IC2S2*.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614–636. https://doi.org/10.1037/a0025255

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning: A Symposium* (pp. 37–46).

Jessop, A., Pine, J., & Gobet, F. (2023). *Chunk-based Incremental Processing and Learning: An integrated theory of word discovery, vocabulary growth, and speed of lexical processing*. PsyArXiv. https://doi.org/10.31234/osf.io/dukpt

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2020). Meaningfulness beats frequency in multiword chunk processing. *Cognitive Science*, *44*(10), e12885. https://doi.org/10.1111/cogs.12885

Kay-Raining Bird, E., & Chapman, R. S. (1998). Partial representations and phonological selectivity in the comprehension of 13- to 16-month-olds. *First Language*, *18*(52), 105–127. https://doi.org/10.1177/014272379801805204

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Mather, E., & Plunkett, K. (2009). Learning words over time: The role of stimulus repetition in mutual exclusivity. *Infancy*, *14*(1), 60–76. https://doi.org/10.1080/15250000802569702

McCauley, S. M., Bannard, C., Theakston, A., Davis, M., Cameron-Faulkner, T., & Ambridge, B. (2021). Multiword units lead to errors of commission in children's spontaneous production: "What corpus data can tell us?*". *Developmental Science*, *24*(6), 1–9. https://doi.org/10.1111/desc.13125

McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, *126*(1), 1–51. https://doi.org/10.1037/rev0000126

Minier, L., Fagot, J., & Rey, A. (2016). The temporal dynamics of regularity extraction in non-human primates. *Cognitive Science*, *40*(4), 1019–1030. https://doi.org/10.1111/cogs.12279

Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, *37*(3), 545–564. https://doi.org/10.1017/S0305000909990511

Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLOS ONE*, *10*(7), e0120644. https://doi.org/10.1371/journal.pone.0120644

New, B., & Pallier, C. (2020). *Lexique 3.83*. http://www.lexique.org

Nickl, A. T., & Bäuml, K.-H. T. (2023). To-be-forgotten information shows more relative forgetting over time than to-be-remembered information. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02330-1

Page, M. P. A., Cumming, N., Norris, D., McNeil, A. M., & Hitch, G. J. (2013). Repetition-spacing and item-overlap effects in the Hebb repetition task. *Journal of Memory and Language*, *69*(4), 506–526. https://doi.org/10.1016/j.jml.2013.07.001

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246–263. https://doi.org/10.1006/jmla.1998.2576

Pinker, S. (1994). *The language instinct: How the mind creates language*. HarperCollins.

Pinto Arata, L., Ordonez Magro, L., Ramisch, C., Grainger, J., & Rey, A. (2024). The dynamics of multiword sequence extraction. *Quarterly Journal of Experimental Psychology*, 1–24. https://doi.org/10.1177/17470218241228548

Rey, A., Bogaerts, L., Tosatto, L., Bonafos, G., Franco, A., & Favre, B. (2020). Detection of regularities in a random environment. *Quarterly Journal of Experimental Psychology*, *73*(12), 1–13. https://doi.org/10.1177/1747021820941356

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Schwab, J. F., & Lew-Williams, C. (2016). Repetition across successive sentences facilitates young children's word learning. *Developmental Psychology*, *52*(6), 879–886. https://doi.org/10.1037/dev0000125

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "Learning" in statistical learning: What does an online measure reveal about the assimilation of visual regularities? *Cognitive Science*, *42*, 692–727. https://doi.org/10.1111/cogs.12556

Sonbul, S., El-Dakhs, D. A. S., Conklin, K., & Carrol, G. (2022). "Bread and butter" or "butter and bread"? Nonnatives' processing of novel lexical patterns in context. *Studies in Second Language Acquisition*, 1–23. https://doi.org/10.1017/S0272263122000237

St-Louis, M.-È., Hughes, R. W., Saint-Aubin, J., & Tremblay, S. (2019). The resilience of verbal sequence learning: Evidence from the Hebb repetition effect. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *45*(1), 17–25. https://doi.org/10.1037/xlm0000566

Swingley, D. (2008). The roots of the early vocabulary in infants' learning from speech. *Current Directions in Psychological Science*, *17*(5), 308–312. https://doi.org/10.1111/j.1467-8721.2008.00596.x

Tamura, N., Castles, A., & Nation, K. (2017). Orthographic learning, fast and slow: Lexical competition effects reveal the time course of word learning in developing readers. *Cognition*, *163*, 93–102. https://doi.org/10.1016/j.cognition.2017.03.002

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, *69*(3), 559–599. https://doi.org/10.1111/lang.12343

Vihman, M. M., DePaolis, R. A., & Keren-Portnoy, T. (2014). The role of production in infant word learning. *Language Learning*, *64*(s2), 121–140. https://doi.org/10.1111/lang.12058

Wang, H. S., Köhler, S., & Batterink, L. J. (2023). Separate but not independent: Behavioral pattern separation and statistical learning are differentially affected by aging. *Cognition*, *239*, 105564. https://doi.org/10.1016/j.cognition.2023.105564

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46–65. https://doi.org/10.1093/applin/aml048

Appendix A. Participants' scores and standard deviations on the LexTALE task for each condition.

Condition	Vocabulary score	*SD*
4F-Condition	87.91%	4.68
7F-Condition	83.04%	6.79
10F-Condition	84.92%	6.74
20F-Condition	88.17%	5.06
30F-Condition	87.03%	5.89
60F-Condition	84.45%	7.44

Appendix B. List of triplets per participant and fillers in all the conditions

Participant	Triplet
1	olive laine flèche
2	olive flèche laine
3	laine olive flèche
4	laine flèche olive
5	flèche olive laine
6	flèche laine olive
7 … 12	melon préfet mythe
13 … 18	évier raisin flacon
19 … 24	lièvre trèfle croûte
25 … 30	grotte truite brosse
31 … 36	reflet mouton poêle
37 … 42	castor brume vapeur
43 … 48	olive laine flèche
49 … 54	melon préfet mythe
55 … 60	évier raisin flacon
61 … 66	lièvre trèfle croûte
67 … 72	grotte truite brosse
73 … 78	reflet mouton poêle
79 … 84	castor brume vapeur

Filler words

abri accent accès acier acte acteur action adieu agence agent agneau aile aise alarme alcool alerte aller allure amant âmes ange angle anneau août appart appel arbre armée arrêt asile assaut aube aura auteur auto avance avions avril bague bail bain baiser baisse balle ballon bande banque barbe baron barre base basket bâton beau beauté belle bête bêtise beurre bible bien bière bijoux billet bisou blague blanc bleu bloc blonde boîte bombe bonne bonté bord bordel bottes bouche boue bouffe bouge boule bourse bouton bruit budget cabane cabine câble cadeau cadre cage caisse camion canal canard cancer canne canon carnet carte carton casier casque casse cave cent centre cercle cesse chaîne chair chaise champ change chant chaos charge charme chasse chat chaud chefs chèque chère chéri chérie cheval choc chou chute cible cieux cigare ciné cirque cité clair classe clef clés client club cochon code cœurs coffre cognac coin colère coma combat commun comte comté congé copain copie copine corde côte coton couche coucou coupe couple coups cour course cousin cran crâne crédit crème crétin crime cris crise croix cuir curé dalle dame danger danse

date décès déesse défaut défi dégâts degrés dehors délire délit demain démon dent départ dépôt désert désir dessin dessus destin détail dette deuil devoir diable dieux dîner dire direct disque doigt dollar donne dose double douche doute dragon drame draps drogue droite droits drôle eaux écart échec écoute écran effet effort église élève empire emploi enfer ennui entrée envers épaule épée époque épouse époux espace espion espoir essai étage étape étoile être étude excuse expert façons faits faveur ferme fêtes feux fiancé fierté fièvre figure file filet films fils fiston flamme fleur fleuve flic flotte foie foire folie folle fonds foot forces forêt forme fort foudre foule four fous foyer frais francs frappe frères frigo froid front fruit fuite fumée fumier fusil futur gaffe gamin gamine gang gants garage garde gare gâteau gauche génie genou geste glace gloire golf gorge gosse goût goutte grâce grain grand grande grève gros groupe guide habits haine hasard hâte haut herbe héros hiver hôte huile humain humeur humour idées idiot idiote image impôts indice infos issue jambe jambes jardin jeudi jeune jeux joie joue jouet joueur juge juif juin jungle jupe jury juste lâche laisse lait lame lampe lance langue lapin larmes leader leçon lèvres lien lieux ligne limite linge lion lires liste livres loge lois long loup loyer lundi lune lutte luxe lycée machin mafia magie maire mairie major malade malin mamie manche mandat manque marche marché mardi marge mariée marque masque masse match membre menace ménage messe mesure métal métier mètres métro mets micro midi miel mieux milieu mille mine minuit miroir mise misère mode modèle moins moitié môme montre moral morale moteur motif moto mouche moyen murs musée nation nature navire neige nerfs neveu niveau noces noir noix nombre noms nord note nuage nuits objet odeur œil œuf œuvre offre oiseau ombre ongles orage orange ordres ordure otage oublie ouest ours outils page pain paire palais panier panne papier paquet parc pardon parent parfum pari parler parole parti passe passé patrie pattes pause pauvre peau péché pêche pensée permis perte peste pétrin phase photos phrase piano pièces piège pierre pile pilote pipe pire piste pitié pizza places plage plaie plans plaque plat plomb pluie plus poche poème poète pognon poids poil poing pointe points poison poker pomme pompe pont porc port portée portes pose poste pote potion pouce poudre poule poulet pouls poupée pour prénom prés presse prêt prêtre preuve prière prime prince prise privé procès profil profit projet promis public puce puits pull quai quart queue race radio rage rançon rang rats

rayon record reçu regard régime région règle reine remède repas repos réseau restes rêve réveil riches rideau rire risque robe robot rocher rois rôle roman rond rose roue rouge routes rues rumeur russe rythme sable sacs saint sainte saison salade salon santé sauce saut saute savoir savon scène séance second secret sein séjour série seul seule sexe shérif siècle siège signal signe singe site sœurs soif soin soirée soirs soldat somme sommet sort sorte sortie souci soupe source souris sous sport stade star statue stop stress studio style stylo succès sucre sueur sujets super survie tabac tache tâche taille talent tante tapis tard tarte tasse taule taux taxi témoin temple tennis tente tenue terme test texte thème thèse ticket tigre tire tireur tiroir titre toile toit tombe tonnes tonton tort touche trace trafic trains trésor trône trou troupe tsar tube tueur tunnel tuyau types union usage usine vache vague valeur valise vallée veille veine vélo vent vente ventre vers vert vertu veste veuve viande vide vierge villa violon virus vision visite vitre vodka vœu voie voile voisin volant voleur vote voyou whisky zéro zone vrai

Filler pseudowords

abréne acogne acrule acun âder adide adit adute aflare afres agarte agnoir ainte alaud âleur alire alme aloux altace ande anels anet anide antise antôt anuche apile apoux appome apsé aqué aquise arais arcime arcin arder arêle aret argu arile arine arplet arron artal arti arut asale asode asor asure atinle atôt augard aure ausi autoce auvé avit avoie balbe balge balve barer barmal barme batel berbe berpe bexe bibe bifait bime blache blaie blait blan blave bleine bligne bligue bloin bloir blone blose bloue blourd blous boce bocle boge bogre boine bolde bonge boque borté bouce boult bralle brappe brare bril brire brix brone brouge buge buin buit bune buve caint cair cait carate carc caruce catail catire catote cauret cavu cédain cepame cèque cercha cerque cerson cetir chague chamir charde charon chassi chaune chefet cheler chens cheur cheve chigue chlouc chobe chol chor chote chotit choui chrite chroid chume chuse cide cieu ciffe cigna cila cile cine cique ciroir cisque claint clapte clar clare clas clat clie clipt cloîte clor cluise coble cocère colmel colte conlé conpé corsai cosin couce coué counai crambe crat crec cres cribe criche crill crobe crogue croile crolle croue crupe crute cuche cuif cumite cuse dade daie daire dalte dalve damade danfum dange daud déceau degite délide delte démote denate dentre dérain derce derler dertes deste dets deur dévede deveu dirie ditige ditrer doble dogne

dondu dornel doste doupe dousus douval drait drie droût drugue druie duri dute duve échace échile écue êder égave eint éjoins êlait élice elsir émin engade ensime entase entat entif êpit éple éques erche éreute erne ertre ésale ésame étade êtele étie étive euve fainé falber fanche fande fanet fanflé farbe farc farie fasque faté fauf fave favec fecore feine fenre fenu ferbe feure feuve ficoce ficule fiège finel flaud fler flerse fles flette flide flie flir floie florve fluie foche folde fope forbi forni forse fouce foude fousse frade frieux frigle frone frore fube fuche fule fune furvis fuste garsu gasse gaune gerte gervec gève gict gide gippe gitre glac gles glie gliffe glin glique glur glyste gnai goir goire gois gomda gonc gose gouée gourte grage graie granal granut grard grare grêche grer grère gretrő greu grez gride grivec groc grode groie grond grort grupe grut guche gumble gure gutte guve haite haites hange hatour hente hèque herde hette heudit heute hiltre hince hoire hombri hôme hontôt hore hôtu hougue huide hule humi humice iant iceaux îcun ideau iesse igre imple ince inge inris insare inssir intrie intute iole ipéth iple ique irare irleme iseau iveau jabe jabon jâche jacun jada jadin jalle japite jarcit jarque jarvu jaut jauve jave jers jerse jexe joite jolée joque jort jouc juie june juse jusé jutand juve juxe kert kigne lage lagne lainté laitôt lâle lanche lanque larais lasque latace laine lave laxe léble léfile lémece lèpe lèpu lère lette leul liba lieune linée lingt lipin lique lisard loce loine loite loqua lorque louque luche luie luve macte maile mance matif mave médin mélé mêlu mèlui mendit menté mertes mesque metsa mevail miant miman minte moif moléce molome mone monner morge morif mosse môvec muche muis mulé munte mûras musé musoin mype myssez naille naise nance nans narte natéme nauffe nelte nertin nieux nigue nime nire noda nogle noil noint nomsi nopore noxe nuce nure ofice ofsa ogne oile oire oisse olde omante ombe ondre onge onstre ontro oques ôques orcin orde orite orlus ornal ornet orral orse oste ouble oudain oude oufage oufre oulais oulu ounite ouve pachet padure panis panque paseau passi pateur pauve pavoi pecran pècun pelte pérée pertif pesque phère phonie phosi phre piale piant pite plache plaze pleité plide ploir plote plou pluc plule pode poge poif poifet poinré pompte porce pordée portin portu poute prade pras preffe premit preper preser preur preuro proce procre proile proir proite prolai prure pruve psat puif punte racare racis ragne raisi raisse raite rande ranu reche refie rèfle règre rèle rêle relle rencan renchi renre renume repis rerile rerise rese resse rétal retave reur reux revate

revote rhague ricit riers rigne rigue rille rils ritel rivete robire rocèle rodi roge roite rong rort rotte roulip ruite rusque sabe sagot salgue salté sarie sarve satron satu scur selte sençon sensin sepoce serte sesin setuce shode sinque siveau sobe soche sodote sofans soile soite solt sona soral sorus sotain souf soumi soumon sounir sousse spel sper spho spile strour suine suir suire sule sunie sunte supe surre suve tabre tade taivec tanse tarche tarme tause taute tecte tèle tenti tervec tese tess têtour teuil teur tiant tiche tiège tieux tince tiplu tipt tise tite tobe toives tole tompe tongle torgue torne touet touge toule toupe toutin toutôt tradit trale trare trât trate traz trent trict trif trile trit triz troce trome tronse trosse trote troule trounu truile trulle trupe tuce tuche tuelle tuille tuis tulque tuste tute tyle uage uant ucun uise unate uque uques urtes uteur vace vadrit vage vaise vala valon vanse vare varte vate vatin vauve vece vecote venne ventin vêpel vesse vets veur vicle vige vigée villon vilume vinde vinge vipole visise vitile vitin voeurs voine voluce vonque voque vore vorie vorve vose vosé voucle voule vouls vrande vroste vuie vule vure wadu zague zitre zoque zote

Appendix C. Parameter estimates for each regression model fitted using beta coefficients as a function of repetition spacing

Model	*a*	*b*	SE *a*	SE *b*	R^2	F	p
Linear	0.01	0.00	0.00	0.00	0.40	2.62	.181
Power	0.02	-0.22	0.00	0.06	0.85	0.63	.036
Exponential	0.01	-0.01	0.13	0.00	0.38	2.43	.194
Logistic	-3.23	0.06	1.51	0.05	0.27	1.48	.290

Note. SE: standard error.

Chapter 4: The lifetime of sequential memory traces in the absence of language

Sequential statistical learning (SSL), the ability to detect sequential regularities in the environment, is a fundamental building block of the development of many cognitive functions, from motor planning to language acquisition. Repetition is a major ingredient in this type of implicit learning, and here we study the effect of repetition spacing on the development of a sequential memory trace. To control for the use of verbal recoding strategies in sequence learning that may influence the learning process, we test a group of Guinea baboons in a Hebb visuo-motor pointing task. This task is a combination of the Serial Reaction Time task and the Hebb repetition paradigm, in which a target sequence is repeated with random sequences inserted between repetitions. In this study, we systematically manipulate the interval between two repetitions of a target sequence by varying the number of interposed random sequences (i.e., repetition spacing). We found that baboons can learn repeated visuo-motor sequences, even when the repetitions are spaced by six random sequences. Our results also suggest that the learning curve best fit a logarithmic function. The present study therefore provides a quantified estimate of the lifetime of a sequential memory trace as a function of repetition spacing and without the benefit of language-related refreshing and recoding strategies[6].

[6] This experimental chapter is an adapted version of the submitted manuscript in thesis format. Ordonez Magro, L., Pinto Arata, L., Fagot, J., Grainger, J., & Rey, A. (Submitted). The lifetime of sequential memory traces in the absence of language.

Introduction

Sequential statistical learning (SSL) is the ability to detect sequential regularities in the environment. It is assumed to be a critical building block for the development of a wide range of cognitive abilities in humans and non-human primates, such as motor planning, spoken and written language acquisition, and social cognition (Frost et al., 2019). Since this phenomenon is involved in the perception and learning of many forms of knowledge, understanding its underlying mechanisms is of utmost importance for advancing general theories of learning in cognitive science (Bogaerts et al., 2020).

The ability to learn repeated and frequent co-occurrences has been observed not only in humans, but also in non-human primates such as tamarins (Hauser et al., 2001), macaques (Wilson et al., 2015), and baboons (Malassis et al., 2018; Minier et al., 2016; Rey et al., 2012, 2019), indicating that regularity extraction can occur in the absence of language and verbal consolidation strategies such as phonological rehearsal and verbal recoding, two verbal processes that facilitate storage in memory (Baddeley, 2003). These studies in non-human primates therefore suggest that sequential learning is probably based on domain-general associative learning mechanisms.

Over the past two decades, in parallel with the growing number of studies aimed at understanding the mechanisms supporting regularity extraction in the environment (see Christiansen, 2019, for a review), several computational models have been developed to account for these processes (e.g., Boucher & Dienes, 2003; Cabiddu et al., 2023; Elman, 1990; Frank et al., 2010; French et al., 2011; Jessop et al., 2023; McCauley & Christiansen, 2019; Monaghan & Christiansen, 2010; Perruchet & Vinter, 1998; Robinet et al., 2011; Servan-Schreiber & Anderson, 1990). For instance, the Simple Recurrent Network (SRN; Elman, 1990), a connectionist neural network trained to predict a given item of a sequence based on previous items in the sequence, suggests that a sequence is learned

through the adjustment of weights in the network after each repetition. As a result, items that are frequently encountered together develop stronger connections with each other. PARSER (Perruchet & Vinter, 1998), another influential model of SSL, assumes that such frequently associated components are grouped together and temporarily encoded as chunks in memory. Importantly, each time a chunk is reencountered, its memory trace is strengthened, and conversely, it progressively decays if it is not encountered again, until it entirely vanishes.

While there is a growing body of evidence that sheds light on the mechanisms that are involved in SSL, evidence regarding the memory limits of SSL is scarce. Indeed, little is known about when exactly a sequential memory trace, that is not repeatedly processed, vanishes. In everyday life, most sequential information that is learned, is not necessarily frequently encountered in the environment. For instance, infants are able to learn multiple words very quickly even when they are not constantly exposed to them (Pinker, 1994; Swingley, 2008). Based on this idea, Page et al. (2013) conducted one of the rare studies that investigate the effect of repetition spacing (i.e., the role of the number of random sequences that intervene between two repetitions of a target sequence) in the learning of sequential material. They used a variant of Hebb's repetition paradigm (Hebb, 1961) in which adult humans are presented with sequences of digits to be recalled immediately. A given sequence (the Hebb sequence) is repeated every third trial and interposed by random (filler) sequences. In their study, Page et al. presented their participants with sequences of syllables in which repetitions were spaced at every 6th, 9th, and 12th trial. They observed that the learning rate for the repeated sequence was substantial and essentially equivalent for all three spacing conditions. These findings

suggest that learning is still possible when repetitions are widely spaced and that memory traces that are not constantly refreshed do not vanish rapidly.

Note, however, that unlike in other species, SSL in humans can be based on the ability to verbally recode non-verbal material (Baddeley, 2003) and to internally repeat sequences, for instance by using the phonological loop (Aboitiz et al., 2010). These verbal consolidation strategies are known to play a key role in maintaining and refreshing information in memory and thus have an undeniable impact on memory performance (Barrouillet & Camos, 2014). Given the close link between memory and language, it is difficult to disentangle memory-specific from language-related mechanisms. One possible way to avoid the presence of any language bias when examining fundamental SSL mechanisms is to test species that lack language. Thus, in the present study, we tested baboons using an adaptation of the Serial Reaction Time task (Nissen & Bullemer, 1987) combined with the advantage of the Hebb repetition paradigm, which allows to control for the spacing between repetitions by inserting a varying number of random filler sequences between each repetition of the target sequence.

In this task, baboons had to touch a red circle that moved on a touchscreen in a fixed sequence of three target locations (selected from a 3x3 matrix of 9 locations). Between two repetitions of the repeated sequence, we inserted either one filler sequence (1F-Condition), three filler sequences (3F-Condition), six filler sequences (6F-Condition), or no filler sequence (0F-Condition). This study involving non-human primates was designed to provide a more accurate estimate of the time a trace can survive in memory without being influenced by language-specific mechanisms, and thus, to inform current SSL models about the properties of natural forgetting in sequence learning.

Experiment

Method

Participants

We tested 25 Guinea baboons (Papio papio, 16 females, age range 2.33 – 24.41 years) living in a social group at the CNRS primate facility in Rousset (France). The baboons were housed in a 700 m^2 outdoor enclosure with access to indoor housing. Water was provided ad libitum during the test, and the monkeys received their normal food ration of fruits every day at 5 PM.

Apparatus

Baboons had free access to fourteen Automated Learning Devices for Monkeys (ALDM, Fagot & Bonté, 2010; Fagot & Paleressompoulle, 2009) equipped with touch screens and a food dispenser. When entering the ALDM test box, baboons were identified by microchips implanted in each arm. The system saved the last trial the baboon had achieved before leaving the box, allowing them to continue the task later where it had stopped. The experiment was controlled by E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). All the baboons were familiar with touch screen experimentation.

Materials and procedure

To begin a trial, baboons had to press the yellow fixation cross centered at the bottom of the screen. After pressing the cross, they saw a black screen that was divided into an invisible matrix of (3x3) cells, each containing a white cross in their center (see Figure 1).

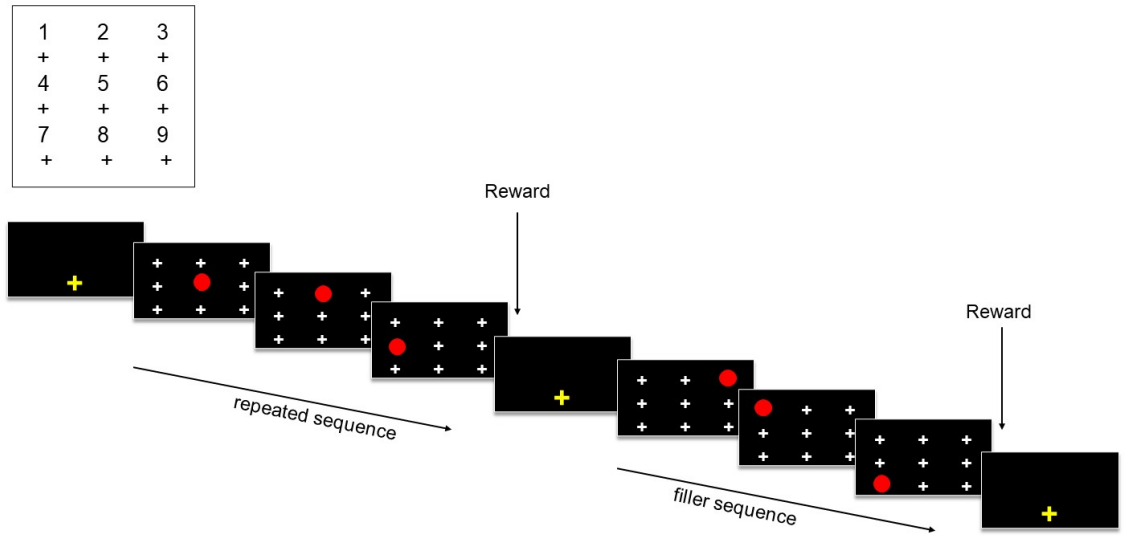


Figure 1. Representation of the touch screen with the nine locations (above) and example of the experimental display and stimuli presentation for repeated and random (filler) sequences in the 1F-Condition (below).

In the task, baboons must touch sequentially one red circle that moved on the screen in sequences of three target locations. When the baboon touched the first target position, it disappeared and was replaced by the white cross. The red circle then appeared at the second position and had again to be touched before being presented with the last position of the sequence, where a last touch was required. Reward (grains of dry wheat) was delivered at the end of each sequence of three correct touches. In case of an error (i.e., the participant touched another location than the target one or failed to touch the screen within a time period of 5,000 ms), a green screen was displayed for 3,000 ms as a marker of failure (see the OSF repository for a video of the procedure).

The task began with a familiarization phase during which baboons were presented with random sequences of three positions. The test phase began when the baboon achieved a performance higher than 80% correct within three consecutive blocks of 100 trials. Response times (RTs) between the appearance of the red circle and the

participant's touch of each of the three positions for each sequence were recorded and served as the main dependent variable. The task was composed of four conditions which differed in the number of filler random sequences introduced between the repeated target sequence. Participants first carried out the 1F-Condition, followed by the 3F-Condition, then the 6F-Condition, and the 0F-Condition. To be presented with a total of 1,000 repetitions of the repeated sequence in each condition, baboons performed 10 blocks of 200 trials in the 1F-Condition, 10 blocks of 400 trials in the 3F-Condition, 10 blocks of 700 trials in the 6F-Condition, and 10 blocks of 100 trials in the 0F-Condition. To avoid learning effects across conditions, baboons were presented with a different repeated sequence in each condition (see Table 1 for an example). Note that repeated and filler sequences were non-overlapping sequences, because the two sequence types used different locations (i.e., when the repeated sequence was composed of the locations 5-2-4, these locations were never used in the filler sequences; see also Ordonez Magro et al., 2022).

In this paradigm, learning is measured by comparing the RTs between the three positions of the sequence. In a repeating three-position sequence such as 5-2-4, location 5 is always followed by location 2 and location 2 is always followed by location 4. Thus, location 5 (Position 1) should remain unpredictable over trials due to the previously presented random sequences. Location 2 (Position 2) should benefit from the systematic presence of 5 presented just before, while location 4 (Position 3) should further benefit from the cumulative contextual information provided by the two preceding positions (Elman, 1990; Minier et al., 2015). If learning occurs, we should observe a predictability effect (Pinto Arata et al., 2024; Rey et al., 2020), reflected in faster RTs for Position 2 over

Position 1 and even faster RTs for Position 3 compared to 2 (Pinto Arata et al., 2024; Rey et al., 2020).

Design of the sequences

To control for the motor difficulty of the transitions between positions of the triplet sequence to be produced, we used data from a prior study conducted on thirteen baboons. In this study, baboons had to produce 1,000 trials each composed of a random sequence of six positions. Based on these random trials, a baseline measure was computed for all possible transitions from one location to another by calculating the mean RTs for each transition (e.g., from position 1 to 9) over all 13 baboons, resulting in a 9x9 matrix of mean RTs (see Appendix A). This baseline measure corresponds to the mean time it takes the baboons to move their hand from one location to another, when the transition is unpredictable (i.e., random). It thus provides a good estimate of the motor difficulty of moving from one position to another.

Importantly, to accurately assess learning, the mean RTs of the first and the second transition of the repeated sequences must be similar before any exposition to the repeated sequence (i.e., the first Transition Time, TT1, from Position 1 to Position 2, and the second Transition Time, TT2, from Position 2 to 3). Therefore, using these baseline measures, we computed TT1 and TT2 for all possible 504 triplets and we selected a set of four triplets (serving as repeated sequences) with the smallest difference between TT1 and TT2 (see Appendix B). For each monkey, each of the four triplets was used in random order in each of the four spacing conditions.

Table 1. Example of sequences presented in the four conditions for a given participant. Digits correspond to the nine screen locations. Repeated sequences are bolded.

Participant	1F-Condition	3F-Condition	6F-Condition	0F-Condition
Subject 1	7-8-1	3-9-6	4-9-2	**2-7-6**
	6-9-2	9-6-7	1-8-5	**2-7-6**
	4-5-1	6-9-8	9-1-8	**2-7-6**
	6-9-2	**5-2-4**	5-2-8	**2-7-6**
	…	3-1-7	5-8-9	…
		3-8-6	8-2-4	
		7-9-1	**7-6-3**	
		5-2-4	9-2-1	
		…	2-5-1	
			5-1-8	
			8-4-9	
			4-5-2	
			1-4-8	
			7-6-3	
			…	

Results

To ensure that all the spacing conditions were comparable, the analysis focused only on the baboons that completed all the conditions. Based on this criterion, we retained the twelve baboons (7 females, age range 4.33 - 22.66 years) that completed all four conditions. Baboons obtained a mean accuracy level of 97.34% in the 1F-Condition, of 97.86% in the 3F-Condition, of 97.62% in the 6F-Condition, and of 96.76% in the 0F-Condition. Incorrect trials were removed from the data set. RTs for each of the three possible positions and for the 1,000 trials were divided into 10 blocks of 100 trials. We normalized the data by removing RTs greater than 800 ms and then performed a recursive trimming procedure to exclude RTs greater than 2 standard deviations from the

mean for each of the three possible positions in a block for each baboon (13.34%, 14.06%, 13.79%, and 13.24% of the data for conditions 1F, 3F, 6F, and 0F)[7].

To measure learning of the repeated sequence, we looked at the evolution of RTs at each of the three positions in the sequence. To adjust for the differences that may appear between positions at the very start of the experiment, we systematically calculated the difference between mean RTs obtained at each block and that obtained at the first block, for every monkey and every position. Data analysis was performed with the R software (version 4.4.1) using linear mixed-effects models (LMEs) fitted with the lme4 package (version 1.1-35.5; Bates et al., 2015). Separate models were fitted for each spacing condition. Each model included Position (1 to 3), Block (1 to 10) and their two-way interaction as fixed effects, and participants as random effects. We used the maximal random-effect structure that allowed convergence. Position was coded using Helmert contrasts (i.e., P1: 0.7 0.0; P2: -0.3 0.5; P3: -0.3 -0.5) to compare Position 1 with both Positions 2 and 3, simultaneously, and Position 2 with Position 3. Block was mean centered to reduce collinearity. RTs were log-transformed prior to each analysis to reduce skewing. Fixed effects were deemed reliable if |t| > 1.96 (Baayen, 2008). The results of the LMEs are summarized in Table 2.

Table 2. Summary of each mixed model for all spacing conditions.

	0F			1F		
Predictors	β	95% CI	*t*	β	95% CI	*t*
Intercept	-53.34	[-61.46, -45.21]	**-12.91**	-35.94	[-46.07, -25.82]	**-6.98**
P1 vs P2-P3	35.60	[21.98, 49.22]	**5.14**	35.20	[24.98, 45.42]	**6.77**
P2 vs P3	38.83	[10.63, 67.03]	**2.71**	33.22	[15.76, 50.68]	**3.74**
Block	-8.14	[-9.04, -7.23]	**-17.70**	-7.11	[-8.65, -5.58]	**-9.11**
P1 vs P2-P3:Block	6.52	[4.60, 8.43]	**6.68**	5.63	[4.11, 7.15]	**7.27**
P2 vs P3:Block	4.86	[2.65, 7.08]	**4.32**	2.42	[0.66, 4.18]	**2.70**

[7] Inspection of the response times distribution revealed that a majority of responses were produced around 500ms. A smaller group of RTs appeared around 1,000 ms and was likely due to situations in which baboon's response was not recorded by the computer, because their hands were dirty. In this situation, they had to touch the screen again, and longer RTs were recorded (that are on average twice longer compared to the first responses). This is why we have adopted this two-step trimming procedure.

Table 2 (continued)

	3F			6F		
Predictors	β	95% CI	*t*	β	95% CI	*t*
Intercept	-28.41	[-36.88, -19.93]	**-6.59**	-19.91	[-27.82, -12.01]	**-4.96**
P1 vs P2-P3	20.25	[10.72, 29.77]	**4.18**	20.46	[5.01, 35.91]	**2.61**
P2 vs P3	17.88	[1.14, 34.63]	**2.10**	9.77	[-6.23, 25.78]	1.20
Block	-6.14	[-7.45, -4.84]	**-9.26**	-4.30	[-6.16, -2.45]	**-4.56**
P1 vs P2-P3:Block	5.47	[4.19, 6.74]	**8.44**	4.29	[3.14, 5.43]	**7.35**
P2 vs P3:Block	3.66	[2.19, 5.13]	**4.89**	2.53	[1.21, 3.86]	**3.76**

Note. P: position. Effects are considered significant when t > |1.96| and are highlighted in bold

The evolution of the difference between mean RTs for all Blocks and mean RTs for Block 1 for all spacing conditions (i.e., 0F, 1F, 3F, and 6F) and all positions in the repeated sequence (i.e., from Position 1 to 3) is presented in Figure 2. Our results showed a significant effect of Block, with RTs decreasing across blocks in all spacing conditions. Consistent with previous findings, we found that baboons produced faster RTs for predictable positions (i.e., Positions 2 and 3) compared to unpredictable positions (i.e., Position 1) in all spacing conditions. Moreover, the difference in RTs between predictable positions and unpredictable positions increased throughout the blocks, as indicated by the significant positive interaction between Position 1 and the average of Positions 2 and 3, and Block. Our results also showed a processing advantage for items in Position 3 over those in Position 2 and a significant increase of this difference across blocks for all spacing conditions.

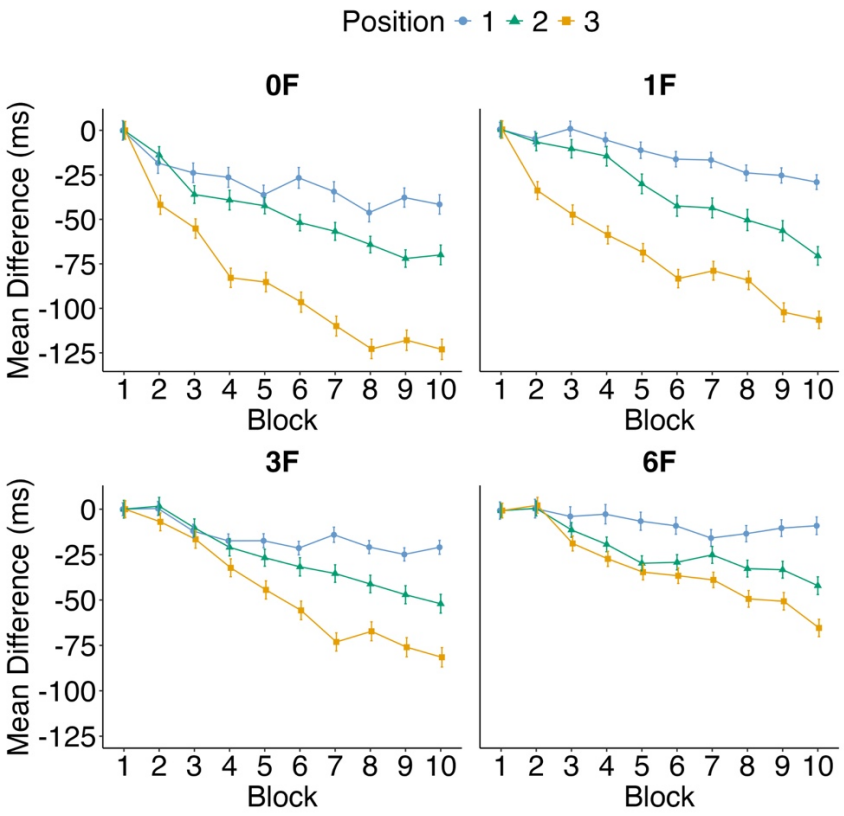


Figure 2. Mean reaction times differences over all baboons between each Block (from Block 1 to 10) and Block 1 for all positions in the sequence (from Position 1 to 3) in all four spacing conditions (0F, 1F, 3F, and 6F). Error bars indicate 95% confidence intervals.

To examine the influence of repetition spacing on sequence learning, we computed a predictability score for each spacing condition by calculating the difference between the RTs for the unpredictable position (i.e., Position 1) and the mean RT for the predictable positions of the triplet for each block (i.e., Positions 2 and 3). Here, a positive score reflects learning of the predictable positions and therefore, learning of the sequence (Pinto Arata et al., 2024). We performed an LME on these predictability scores, using spacing condition and block as fixed effects, along with the interaction between these variables, and by-subject random intercepts, as well as by-subject random slopes for block. Spacing condition was coded using repeated contrast coding to perform pairwise comparisons.

We found higher predictability scores in the 0F-Condition compared to the 1F-Condition, b = -0.04, 95% CI [-0.06, -0.02], t = -4.44, and higher scores in the 1F-Condition compared to the 3F-Condition, b = -0.05, 95% CI [-0.07, -0.03], t = -4.91. There was no difference between the 3F- and 6F-Condition. We also observed a significant effect of Block, b = 0.02, 95% CI [0.01, 0.02], t = 8.24, indicating an increase of predictability scores across blocks.

To obtain an estimate of the time that a sequential memory trace can survive without being reinforced, we examined the evolution of the predictability score over the different repetition spacing. When the predictability score becomes null (i.e., when there is no longer any difference between unpredictable Position 1 and predictable Positions 2 and 3), then we can assume that the spacing between two repetitions is too important to consolidate the sequence's memory trace. We ran a regression analysis on the beta coefficients of each linear regression of the mixed model described above to determine which function best fit the dynamics we observed for the predictability scores (the greater the beta coefficient, the greater the learning of the sequence). The models were compared using AIC, BIC, and the R2 scores (see Table 3). Lower AIC and BIC scores and higher R2 values indicate a better model fit. Results revealed that the evolution of the predictability scores was better captured by a logarithmic function more than by a power, exponential or linear function. Figure 3 shows the results of the LME for the predictability score across the four spacing conditions together with the logarithmic function for the distribution of the beta coefficients.

Table 3. Model comparison with parameter estimates for each model

Model	*a*	*b*	AIC	BIC	R^2
Logarithmic	0.015	0.001	-39.542	-41.383	0.937
Power	0.015	-0.071	-38.537	-40.378	0.901
Exponential	0.019	-0.088	-36.292	-38.133	0.889
Linear	0.019	-0.001	-35.487	-37.328	0.825

Note. AIC: Akaike's information criterion; BIC: Bayesian information criterion.

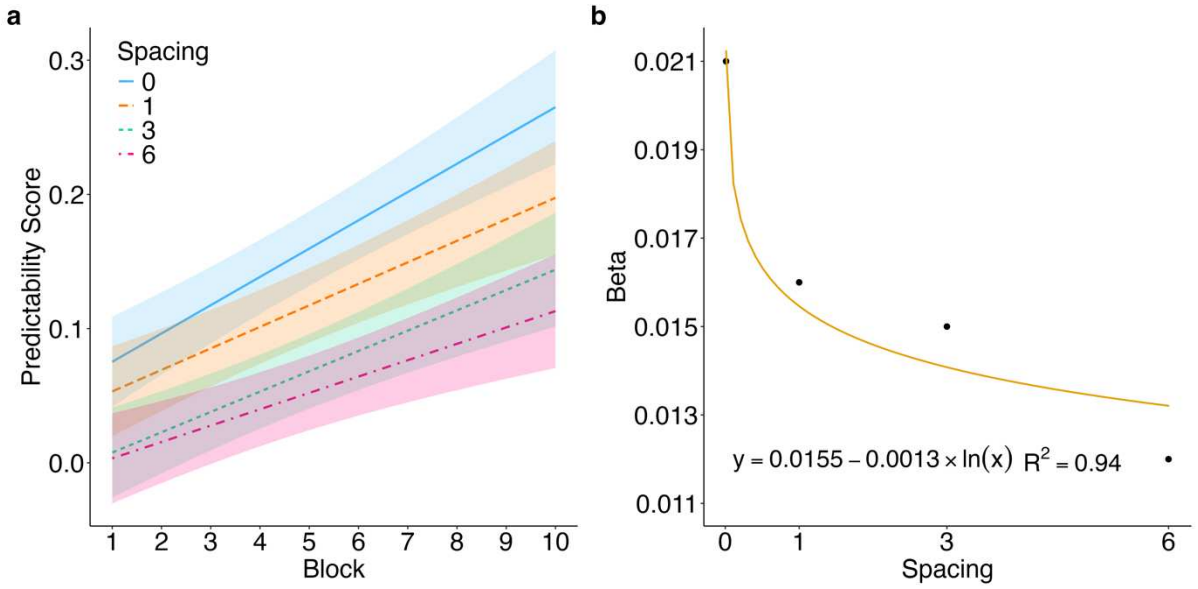


Figure 3. Panel a: Comparison of predictability scores across blocks for repeated sequences in all the spacing conditions. Continuous lines represent the best linear fit and color-shaded areas indicate 95% confidence intervals around linear regression lines. Panel b: Beta coefficients for all four spacing conditions. The continuous line represents the best fit to the data (i.e., a logarithmic function). Error bars indicate 95% confidence intervals for each beta coefficient based on the linear mixed model used to compare spacing conditions.

Discussion

In the present study, we used a combination of the Serial Reaction Time task (Nissen & Bullemer, 1987) and of the Hebb repetition paradigm (Hebb, 1961), in which a variable number of random sequences were inserted between two repetitions of a target sequence

to test the limit at which a sequential memory trace vanishes. To do this, we systematically manipulated the spacing between two repetitions of the target sequence by inserting zero to six random filler sequences. Consistent with previous studies (Malassis et al., 2018; Minier et al., 2016; Rey et al., 2019, 2022), we observed that baboons were able to learn sequences of three locations, as indicated by the decrease in RTs for the predictable second and third positions in the sequence across repetitions, even for the largest spacing. The observation of faster RTs for the third compared to the second position is also consistent with previous findings showing that the third position benefits from stronger contextual information during learning compared to the second (Elman, 1990). This advantage suggests that the occurrence of the third position is progressively predicted by the co-occurrence of the first and second positions.

To our knowledge, the present study is the first to provide quantitative evidence on the role of repetition spacing in the formation of stable sequential memory traces by controlling for the influence of verbal consolidation strategies and by testing a non-human primate species (here, Guinea baboons *papio papio*). More importantly, our results showed that learning of the sequence (measured by a predictability score) decreased as spacing increased, and that the relationship between spacing and sequence learning best fitted a logarithmic function. This result suggests that the memory trace of a sequence survives by following the non-linear dynamics of a logarithmic function, and that even if we were to further increase the spacing between two repetitions of the target sequence, we should still observe learning of the sequence (manifested here by an increasing difference in response time between predictable and non-predictable positions). This has important implications for current models of statistical learning.

The idea that sequence learning is based on the creation and consolidation of chunks is supported by several influential chunking models (Frank et al., 2010; French et al.,

2011; McCauley & Christiansen, 2019; Perruchet & Vinter, 1998; Robinet et al., 2011) and, to our knowledge, the results observed in the current study provide the first quantitative estimate of the lifetime of chunk representations. For one of the most influential models in this field, i.e., the Parser model, each time a new chunk is created in perceptual memory, this new unit receives a certain weight that materializes the strength of its trace in memory. This weight then decreases linearly if the chunk is not encountered and processed again. So, after a relatively short time (20 processing cycles), the memory trace of this chunk may disappear if it has not been processed again. Our data suggest that each new sequence that is transiently memorized seems to survive much longer than predicted by the Parser model. It would thus be important to see to what extent a much greater spacing of a repeated sequence (e.g., 20) would lead to learning of this sequence. Unfortunately, this type of experiment is too costly with the paradigm used here with baboons, because as Figure 2 shows, we only start to observe a difference between predictable and non-predictable positions from the third Block onwards in the maximum spacing condition (i.e., 6F). However, recent human data obtained in a Hebb lexical decision task (i.e., Pinto Arata et al., 2024) indicate that learning is possible after a spacing of 60 (Pinto Arata, Ramisch et al., submitted). Whereas in the present study a logarithmic function provided the best fit to our results, Pinto Arata, Ramisch et al. found that learning was best fitted by a power law. This is not surprising, since both functions have been shown to be highly correlated across different learning tasks (Radvansky et al., 2024). Indeed, if one fits well, the other is likely to do so as well. This result validates the idea of a very slow disappearance of the sequence's memory trace, which appears to be captured by the dynamics of either a logarithmic or a power function.

Similarly, McCauley and Christiansen's (2019) recent model also assumes a mechanism for chunk creation as words and word sequences are processed, but unlike Parser, it

assumes that these new chunks are never forgotten once they have been created. Even if there is no empirical data to validate this strong hypothesis, we can nevertheless consider that if the dynamics of forgetting the memory trace of a sequence follows a logarithmic function, then the trace decreases asymptotically and, as a result, its survival time is potentially very long. Our data therefore seem to be fairly consistent with the hypothesis formulated by this model concerning the extremely long survival of chunks that are created over time and exposure to language sequences.

The idea that these sequential memories have a very long lifetime can also certainly help to resolve the question of how infants are able to learn relatively infrequent words (Pinker, 1994; Swingley, 2008). Indeed, representations of words, even when being encountered relatively infrequently, can remain in long-term memory for years once they are part of the mental lexicon. This would be linked to the undoubtedly general property of non-linear decay of sequential memory traces, which follow a logarithmic function. The present study suggests that this general property is probably also present in other primate species, such as the Guinea baboon.

Beyond the long survival of sequential memory traces, it should be noted that in humans, with the development of motor control of the oral-phonatory apparatus during the first year of life, young children are then able to generate verbal sequences themselves, producing repetitions that can compensate for the poverty of surrounding linguistic stimuli. By means of inner language, they can thus self-refresh verbal sequential memory traces, increasing the strength of these memory trace and the efficiency of the

processing of this information within their language perception and production systems (Chater et al., 2016).

Conclusion

The present study demonstrates that a non-human primate species lacking language can maintain and consolidate a sequence of elements even when they are interleaved with a relatively large number of random sequences. These findings suggest that regularity extraction is possible in the absence of language and may rely on language-independent associative learning mechanisms alongside language-related refreshing mechanisms. Most importantly, the current study is the first to provide an estimate of the natural impact of forgetting on a sequential memory trace that does not benefit from verbal consolidation strategies. It indicates that the memory traces of sequential information certainly have a much longer lifetime than previously assumed.

Ethical approval and Informed Consent

This research adhered to the applicable French rules for ethical treatment of research animals and received ethical approval from the French Ministery of Education (approval APAFIS#2717-2015111708173794 10 v3).

Open Practices Statements

We report all data exclusions, all manipulations, and all measures and statistical tools in the study in the Method section. All data and analysis code are available online at https://osf.io/vdyt7/?view_only=a2a7a16cb7694e1da50ed3a3aa563673.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research was supported by the Convergence Institute ILCB (ANR-16-CONV-0002), the COMPO ANR project (#ANR-23-CE23-0031), the HEBBIAN ANR project (#ANR-23-CE28-0008), and the ERC advanced grant (#POP-R 742141).

References

Aboitiz, F., Aboitiz, S., & García, R. R. (2010). The phonological loop : A key innovation in human evolution. *Current Anthropology*, 51(S1), S55-S65. https://doi.org/10.1086/650525

Baayen, R. H. (2008). *Analyzing linguistic data : A practical introduction to statistics using R*. Cambridge University Press.

Baddeley, A. (2003). Working memory and language : An overview. *Journal of Communication Disorders*, *36*(3), 189-208. https://doi.org/10.1016/S0021-9924(03)00019-4

Barrouillet, P., & Camos, V. (2014). *Working memory : Loss and reconstruction*. Psychology Press.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bogaerts, L., Frost, R., & Christiansen, M. H. (2020). Integrating statistical learning into cognitive science. *Journal of Memory and Language*, *115*, 104167. https://doi.org/10.1016/j.jml.2020.104167

Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science*, *27*(6), 807-842. https://doi.org/10.1207/s15516709cog2706_1

Cabiddu, F., Bott, L., Jones, G., & Gambi, C. (2023). CLASSIC utterance boundary : A chunking-based model of early naturalistic word segmentation. *Language Learning*, 1-34. https://doi.org/10.1111/lang.12559

Chater, N., McCauley, S. M., & Christiansen, M. H. (2016). Language as skill : Intertwining comprehension and production. *Journal of Memory and Language*, *89*, 244-254. https://doi.org/10.1016/j.jml.2015.11.004

Christiansen, M. H. (2019). Implicit statistical learning : A tale of two literatures. *Topics in Cognitive Science*, 1-14. https://doi.org/10.1111/tops.12332

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179-211. https://doi.org/10.1016/0364-0213(90)90002-E

Fagot, J., & Bonté, E. (2010). Automated testing of cognitive performance in monkeys : Use of a battery of computerized test systems by a troop of semi-free-ranging baboons (Papio papio). *Behavior Research Methods*, *42*(2), 507-516. https://doi.org/10.3758/BRM.42.2.507

Fagot, J., & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. *Behavior Research Methods*, *41*(2), 396-404. https://doi.org/10.3758/BRM.41.2.396

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107-125. https://doi.org/10.1016/j.cognition.2010.07.005

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX : A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614-636. https://doi.org/10.1037/a0025255

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research : A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128-1153. https://doi.org/10.1037/bul0000210

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate : Statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53-B64. https://doi.org/10.1016/S0010-0277(00)00132-3

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Éd.), *Brain mechanisms and learning : A Symposium* (p. 37-46).

Jessop, A., Pine, J., & Gobet, F. (2023). *Chunk-based incremental processing and learning : An integrated theory of word discovery, vocabulary growth, and speed of lexical processing*. PsyArXiv. https://doi.org/10.31234/osf.io/dukpt

Malassis, R., Rey, A., & Fagot, J. (2018). Non-adjacent dependencies processing in human and non-human primates. *Cognitive Science*, *42*(5), 1677-1699. https://doi.org/10.1111/cogs.12617

McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use : A cross-linguistic model of child language development. *Psychological Review*, *126*(1), 1-51. https://doi.org/10.1037/rev0000126

Minier, L., Fagot, J., & Rey, A. (2016). The temporal dynamics of regularity extraction in non-human primates. *Cognitive Science*, *40*(4), 1019-1030. https://doi.org/10.1111/cogs.12279

Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound : Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, *37*(3), 545-564. https://doi.org/10.1017/S0305000909990511

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning : Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1-32. https://doi.org/10.1016/0010-0285(87)90002-8

Ordonez Magro, L., Fagot, J., Grainger, J., & Rey, A. (2022). On the role of interference in sequence learning in Guinea baboons (Papio papio). *Learning & Behavior*. https://doi.org/10.3758/s13420-022-00537-1

Page, M. P. A., Cumming, N., Norris, D., McNeil, A. M., & Hitch, G. J. (2013). Repetition-spacing and item-overlap effects in the Hebb repetition task. *Journal of Memory and Language*, *69*(4), 506-526. https://doi.org/10.1016/j.jml.2013.07.001

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246-263. https://doi.org/10.1006/jmla.1998.2576

Pinker, S. (1994). *The language instinct : How the mind creates language*. HarperCollins.

Pinto Arata, L., Ordonez Magro, L., Ramisch, C., Grainger, J., & Rey, A. (2024). The dynamics of multiword sequence extraction. *Quarterly Journal of Experimental Psychology*, 1-24. https://doi.org/10.1177/17470218241228548

Radvansky, G. A., Parra, D., & Doolen, A. C. (2024). Memory from nonsense syllables to novels : A survey of retention. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-024-02514-3

Rey, A., Bogaerts, L., Tosatto, L., Bonafos, G., Franco, A., & Favre, B. (2020). Detection of regularities in a random environment. *Quarterly Journal of Experimental Psychology*, *73*(12), 2106-2118. https://doi.org/10.1177/1747021820941356

Rey, A., Fagot, J., Mathy, F., Lazartigues, L., Tosatto, L., Bonafos, G., Freyermuth, J., & Lavigne, F. (2022). Learning higher-order transitional probabilities in nonhuman primates. *Cognitive Science*, *46*(4). https://doi.org/10.1111/cogs.13121

Rey, A., Minier, L., Malassis, R., Bogaerts, L., & Fagot, J. (2019). Regularity extraction across species : Associative learning mechanisms shared by human and non-human primates. *Topics in Cognitive Science*, *11*(3), 573-586. https://doi.org/10.1111/tops.12343

Robinet, V., Lemaire, B., & Gordon, M. B. (2011). MDLChunker : A MDL-Based cognitive model of inductive learning. *Cognitive Science*, *35*, 1352-1389. https://doi.org/10.1111/j.1551-6709.2011.01188.x

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 592-608. https://doi.org/10.1037/0278-7393.16.4.592

Swingley, D. (2008). The roots of the early vocabulary in infants' learning from speech. *Current Directions in Psychological Science*, *17*(5), 308-312. https://doi.org/10.1111/j.1467-8721.2008.00596.x

Wilson, B., Smith, K., & Petkov, C. I. (2015). Mixed-complexity artificial grammar learning in humans and macaque monkeys : Evaluating learning strategies. *European Journal of Neuroscience*, *41*(5), 568-578. https://doi.org/10.1111/ejn.12834

Appendix A. Mean response times over a group of 13 baboons for each of the 72 possible transitions calculated from 1,000 random trials.

1st position in Transition	2nd position in Transition								
	1	2	3	4	5	6	7	8	9
1		426	421	438	365	360	447	359	371
2	506		457	411	377	393	391	365	393
3	502	435		443	368	353	439	372	365
4	486	423	448		366	374	434	339	358
5	485	408	378	444		345	449	392	380
6	477	383	379	426	344		448	384	418
7	472	424	435	423	370	381		374	371
8	445	388	401	396	342	367	443		396
9	487	403	410	425	334	361	437	362	

Appendix B. The four selected repeated sequences and their respective mean TT1 and TT2 values.

Triplet	TT1	TT2
524	408	411
692	418	403
276	391	381
763	381	379

Note. TT = transition time.

Chapter 5: Chunking mechanisms in language and other domains

The ability to track the statistical structure of language and more broadly, of our environment, is a key feature of our cognitive system. This process, known as statistical learning, is thought to rely on associative mechanisms, and notably chunking. In this review, we summarize recent empirical work on three main phenomena that have been consistently reported in the literature about chunking mechanisms: predictability effects, repetition spacing effects, and chunk size limits. To illustrate the generality and robustness of these phenomena, we show that they have been observed for the processing of both linguistic and visuo-motor sequences, in human and non-human primate studies. We discuss how current chunk-based models of statistical learning can account for these effects and highlight some of their limitations. Finally, we argue that recent neurocomputational models based on associative and Hebbian learning may provide new theoretical approaches to describe and better understand the nature of chunking mechanisms[8].

[8] This chapter is an adapted version of the manuscript in thesis format. The published version can be found here: Pinto Arata, L., Tosatto, L., & Rey, A. (2024). Chunking mechanisms in language and other domains. *L'Année Psychologique/Topics in Cognitive Psychology*.

Introduction

Our world is full of statistical regularities that humans can exploit to build coherent representations of their environment. This ability to detect and extract these regularities, known as statistical learning, is considered to be a key feature of our cognitive systems. Such statistical sensitivity is thought to rely on associative learning mechanisms, and one in particular, *chunking* (Isbilen et al., 2020; Perruchet & Vinter, 2002; Rey et al., 2019).

Chunking refers to the process of combining and grouping co-occurring sequences of items into single processing units called *chunks* (Gobet et al., 2001, 2016). This mechanism has been studied in a wide range of domains, such as language learning (e.g., Isbilen & Christiansen, 2018; Perruchet et al., 2014; Pulido, 2021), expertise (e.g., Chase & Simon, 1973; Gobet & Simon, 1998; Hu & Ericsson, 2012), social cognition (e.g., Basyouni & Parkinson, 2022; Stahl & Feigenson, 2014), visual sequence learning (e.g., Avarguès-Weber et al., 2020; Daltrozzo et al., 2017; Emerson & Conway, 2023), and motor sequence learning (e.g., Minier et al., 2016; Rey et al., 2019; Tosatto, Fagot, et al., 2022). In addition, chunking is at the heart of many computational models of statistical learning (e.g., Fonollosa et al., 2015; Gobet et al., 2001; Mareschal & French, 2017; Orbán et al., 2008; Robinet et al., 2011), particularly in language processing at various levels of granularity (e.g., Cabiddu et al., 2023; French et al., 2011; McCauley & Christiansen, 2019a; Perruchet & Vinter, 1998).

The main idea behind chunking is that grouping items together into larger units will lead to the creation of a single memory trace, resulting in a unified representation of previously separate pieces of information. This allows our cognitive system, firstly, to compress information to overcome memory limitations and increase short-term memory capacity (Cowan, 2001; Mathy & Feldman, 2012; Miller, 1956), and secondly, to retrieve

and use encoded information faster, thus facilitating several cognitive processes (Christiansen & Chater, 2016; Fonollosa et al., 2015).

Deliberate versus automatic chunking

According to Gobet et al. (2001), chunking can be classified as either *deliberate* or *automatic*, depending on how and when it occurs. On the one hand, deliberate chunking is thought to be controlled, goal-oriented, conscious and explicit (Mathy et al., 2016). For instance, one might use a strategy to remember a phone number by grouping adjacent digits into bundles (i.e., 06.19.39.19.45) or by recoding it from information stored in long-term memory (i.e., 06 + beginning and end of WWII). Automatic chunking, on the other hand, is defined as a non-controlled, unconscious and implicit continuous process in which co-occurring items are grouped together merely by repetition, without any intention to learn (Gobet, 2017; Perruchet & Pacton, 2006). While the former is usually related to short-term memory, the latter deals with processes occurring in long-term memory (Gobet et al., 2016).

Here we focus on automatic chunking as it is considered to be the main mechanism involved in statistical learning. Indeed, statistical learning has been described as occurring automatically, incidentally and without awareness (Batterink et al., 2017, 2019). Note that statistical learning is also referred to as implicit statistical learning in the literature (see Christiansen, 2019 for a review).

Automatic chunking: a fundamental mechanism in sequence learning

As noted previously, chunking has been widely recognized as a domain-general learning mechanism in a number of different areas. In particular, automatic chunking is thought to be essential for language learning, as children do not initially store linguistic information by explicit and rote learning, but learn implicitly (Gobet, 2017). This type of chunking is also crucial for sequence processing in language at different levels. For

instance, at the word-level, it has been suggested that humans group letters together to develop larger processing units, such as graphemes and syllable constituents like onsets or rimes (Brand et al., 2007; LaBerge & Samuels, 1974; Perry, 2023; Rey et al., 1998, 2000; Rey & Schiller, 2005; Ziegler & Goswami, 2005). Similarly, at the phrase-level, it has been shown that infants and adults tend to group words together into multiword units and use them during language comprehension (e.g., Arnon & Clark, 2011; Carrol & Conklin, 2020; Havron & Arnon, 2020) and production (e.g., Arnon & Priva, 2013; Bannard & Matthews, 2008).

However, when it comes to studying automatic chunking, individuals' preexisting linguistic knowledge and expectations of regularities in language make it difficult to isolate the influence of explicit and implicit processes on learning. It has been suggested that one way to overcome this issue is to study this mechanism in species that do not use language to avoid any explicit recoding process (Minier et al., 2016), thus allowing these factors to be disentangled by directly comparing human and animal performance (Tosatto, Fagot, & Rey, 2023b). Indeed, humans and some other species, such as non-human primates, are believed to share common associative learning mechanisms necessary for sequence learning (Malassis et al., 2018; Rey et al., 2019, 2022).

Even though the notions of *chunk* and *chunking* have been well described theoretically (e.g., Gobet et al., 2001, 2016), and demonstrated experimentally (e.g., Emerson & Conway, 2023; Perruchet, 2019; Slone & Johnson, 2018), the exact nature of the associative mechanisms involved has remained elusive. In particular, it is unclear how the processing of regular sequences longer than two items occurs at the neurobiological level, and how the development of a chunk translates into neurobiological mechanisms. While some studies have investigated which brain regions are involved in statistical learning (e.g., Batterink et al., 2019; Conway, 2020; Conway et al., 2020; R. Frost et al., 2015), these

fundamental questions have been surprisingly overlooked in the literature. We therefore still need to understand how we segment a repetitively processed sequence of more than two elements into chunks, and the nature of the associative mechanisms that enable such segmentation.

In the present article, we review three recent key empirical phenomena that are present in the sequential learning of linguistic and non-linguistic regularities: predictability effect, repetition-spacing and chunk size. Our main goal is to further characterize automatic chunking and its fundamental associative learning properties. To this end, we examine recent studies investigating chunking in both multiword sequence extraction in humans and visuo-motor sequence learning involving non-human primates (and more specifically, Guinea baboons, *Papio papio*). For each empirical phenomenon, we present results from linguistic tasks and baboon data, and then we discuss the extent to which current chunk-based models of statistical learning can effectively account for these phenomena.

Predictability effects

Predictability effects in language

Many recent studies have focused on learning sequences composed of three ABC elements. By presenting ABC repeatedly and using experimental protocols that allow implicit statistical learning to be studied in real time (and not following the offline protocols that first appeared in this literature, such as Saffran et al.'s, 1996), it is possible to study the evolution of learning on each element of the ABC sequence. While a simple associative model predicts that the association between A and B will develop as quickly as the association between B and C, numerous studies show that BC is learned more quickly than AB. One way of explaining this result is to consider that, unlike B, C benefits from richer contextual information that include not only B, but also A. Indeed, both A and

B, which systematically appear before C, predict C, while only A predicts B. This processing advantage for C due to the richer context is also known as the *predictability effect* (Pinto Arata, Ordonez Magro, et al., in press).

In the field of language processing, a number of recent studies have investigated these predictability effects with implicit exposure protocols to repetitions of linguistic regularities composed of three ABC elements and real-time measures of learning of these regularities. For instance, Rey et al. (2020) used a letter-naming task where participants had to read aloud letters presented one by one in the center of the screen. Unbeknownst to the participants, a triplet of three letters (e.g., AOU) was systematically repeated in the same order, and 3 to 5 filler letters were inserted between two repetitions of the target triplet. They observed that participants named the second and third letters of the triplet increasingly faster as the task progressed, and that these letters gradually became more predictable for the participants. Most notably, they found faster reaction times (RTs) for the third letter over the second, indicating that the last item of the triplet benefited from richer contextual information (i.e., AO in AOU provides larger contextual information to predict U).

More recently, Pinto Arata, Ordonez Magro et al. (in press) adapted Rey et al's (2020) into a lexical decision task to study multi-word sequence acquisition across four different conditions. Here, the target triplet of letters was replaced by a triplet composed of unrelated words (e.g., *mule proie noeud* – "mule prey knot"), pseudowords (e.g., *fitre plou boge*), semantically related words (e.g., *banane cerise fraise* – "banana cherry strawberry") and idioms (e.g., *faire profil bas* – "lie low"), respectively in the four experimental conditions. They observed faster processing for the second and third items of the triplet throughout the repetitions. Furthermore, the learning dynamic also varied as a function of the nature of the triplet, with semantically related words and idioms being learned

faster. However, no processing advantage for the third position over the second was found. This was explained by the nature of the task as lexical decision demands longer processing time and produces a greater attentional load than naming tasks, thus hindering the establishment of an association between the first and third items of the triplet.

These empirical results therefore show that, in the language domain, contextual information influences sequence processing. The richer and more predictable the context, the more likely it is that associations will be established between the items of a given sequence. However, the nature of the task may favor or hinder the creation of an association, depending on the cognitive load required to achieve it. Furthermore, results from corpus analyses of child directed speech also show a learning advantage for the last words in an utterance. Indeed, children tend to understand these words better and produce them more often (Braginsky et al., 2019).

Predictability effect in other domains

The effect of predictability – as evidenced by an advantage of the last element(s) in the learning of an ABC sequence – has been observed in nonverbal and nonlinguistic tasks as well, especially in visuo-motor tasks, which allow for testing different species and for evaluating the generality of this learning property. In non-human animals, this effect of predictability has indeed been repeatedly observed with studies involving Guinea baboons. For example, Minier et al. (2016) studied the learning of a repeated sequential triplet ABC and the dynamics of learning for each item composing this triplet. This experiment uses a serial response time (SRT, Nissen & Bullemer, 1987) task in which baboons have to track and touch a moving target on a touch screen. In this task, baboons are placed in front of a touch screen that is divided into nine equidistant locations represented by white crosses on a black background, one of them being replaced by the

target, a red circle (see Figure 1, Panels A and B). When the target is touched, it disappears and it is replaced by the cross. The next cross/location in the sequence is then replaced by the red circle until the end of the sequence. Baboons are exposed to 9-item sequences composed of three independent ABC triplets of fixed locations/items and perform 2,000 trials each, one trial being composed of a random combination of the three triplets. RTs are recorded for each touch of the target throughout the task and allow to observe how each item is learned. Indeed, as each triplet can be positioned first, second or third in the sequence, the appearance of the first item of a triplet is not predictable and no significant decrease in RT should be observed for this item. But the next two items are predictable based on the presentation of the first one. Thus, results show a decrease in RT for the last two items but crucially, this decrease is significantly stronger for the third item than for the second one. This advantage for final over initial items appears very early during learning, and it persists during the whole experiment. More recently, Ordonez Magro et al. (2022) have replicated this effect using a modified version of this experimental task (these results are discussed further in Section 3.2).

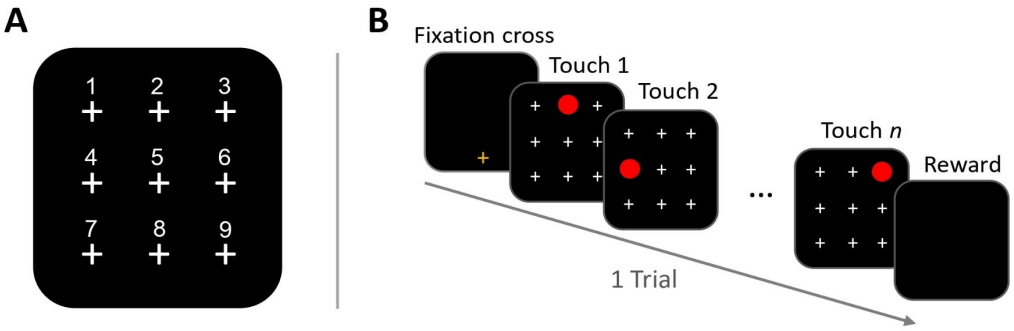


Figure 1. (A) Representation of the touch screen with the nine possible locations. Numbers are shown for illustrative purposes but were not displayed. (B) Example of a single trial. After an initial touch of the fixation cross, the subject had to touch the red dot when it appeared at each location.

Interestingly, similar results are observed with human participants in a strictly identical SRT task (Rey et al., 2019). For both humans and baboons, the advantage for final over initial items in a pattern can be easily interpreted again as the last item in the triplet benefiting from a richer contextual information (i.e., the two previous items) compared to the second item, which only benefits from only one previous item. The similar results obtained in SRT tasks in both human and non-human primates suggest that this beneficial effect of context is a general feature of the chunking mechanisms operating during sequence learning and that it relies on shared fundamental associative mechanisms for both species.

Similarly, in visual statistical learning tasks with human participants, some studies have also isolated a better performance on the third item of a triplet compared to the second item. For instance, Turk-Browne et al. (2005) observe a predictability effect on a visual target detection task. Here, participants are shown four sequential recurring triplets of shapes during an exposition phase. In a subsequent test phase, participants view sequences of 24 items composed of two repetitions of the four triplets in a random order and have to press a key as fast as possible when they detect the target stimulus - one of the shapes, defined at the beginning of each trial. Results show that detection RTs are significantly faster in trials where the target is the third item of the triplet compared to the second item. Likewise, Kim et al. (2009) observe the same results in a similar target detection task and show the long-lasting effects of predictability provided by a richer context, as this advantage of the third item in the triplet over the second is present even when the exposition and test phases are conducted 24 hours apart.

This series of experiments on the extraction of repeated triplets raises the question of the impact of the number of co-occurring elements on this predictability effect. Indeed, what happens if the chain of co-occurring items increases? Would an additional decrease

of RT be observed on a fourth item in a repeated quadruplet? As we will see in Section 4 on chunk size, other experiments involving the learning of longer sequences show that this predictability effect is modulated by sequence size and what appears to be a fundamental limitation in our ability to extract chunks larger than 3 or 4 in size (Tosatto, Fagot, & Rey, 2023b; Tosatto, Fagot, et al., 2022, these studies are further discussed in Section 4).

Taken together, these studies on the implicit learning of triplet sequences that reveal a predictability effect shed light on the architecture of the associations formed between units of a chunk and provide information for constructing computational models of chunking. These theoretical implications will be discussed in Section 5.

Repetition spacing

Repetition spacing in language

It is possible to study the effect of repetition spacing by manipulating the distance between each occurrence of a repeated sequence. These spacing effects thus inform us about the dynamics of the development of a memory trace of this repeated sequence. The question is whether these repetition spacing effects function as simple repetitions of a single piece of information, or whether there are phenomena specific to the repetition of sequences, such as ABC triplets.

Although the effect of repetition spacing on learning has been investigated for over a century (see Cepeda et al., 2006, for a review), only a handful of studies have since focused on its effect on sequence learning. For instance, Melton (1963) used the Hebb repetition task (Hebb, 1961) where participants were presented with sequences of digits for immediate serial recall. Unlike the original task, in which each target sequence (the Hebb sequence) is repeated every third trial and interspersed by random (filler) sequences, the number of sequences between each Hebb sequence varied. Melton (1963) observed that

beyond six filler sequences no learning could be observed. Similarly, Page et al. (2013) presented participants with sequences of words in three different spacing conditions: on every sixth, ninth and twelfth trial. However, in contrast to Melton's (1963) findings, the authors found that learning was substantial and equivalent across all conditions. Furthermore, it has also been shown that sequence learning can occur even when repetition spacing is irregular (St-Louis et al., 2019).

These findings demonstrate that memory traces can survive wide and irregular spacing between repetitions, yet the repetition of sequences of items we encounter in everyday life are usually much more spaced than in current experiments. Based on this observation, Pinto Arata, Ramisch et al. (2024) went beyond these results to investigate the effect of longer spacing on the acquisition of word sequences in real-time. Using the lexical decision task described in Section 2.1., they presented their participants with a triplet of words which repetitions were spaced between: 4, 7, 10, 20, 30, and 60 filler items. In line with previous studies (e.g., Pinto Arata, Ordonez Magro, et al., in press; Rey et al., 2020), they observed faster RTs for the predictable positions (i.e., the 2nd and the 3rd) in the triplet as repetitions progressed (i.e., the predictability effect). It is worth noting that the learning rate differed across different spacing conditions. In fact, the overall decrease in RTs was greater when the spacing was 4 fillers and slower when it was 60 fillers. Furthermore, they also found that RTs for the unpredictable first position decreased with repetition (i.e., the classical repetition effect), but significantly less and at a slower rate than for words appearing in the predictable positions, which processing not only benefits from repetition but also from the richer context, as discussed in Section 2.1. Such findings suggest that although the learning rate is affected by the spacing, sequence learning remains possible even when the repetition spacing between repetitions is relatively large. In addition, these results also indicate that the dynamics of the development of a memory

trace are differentially affected by the repetition and predictability of the items in the repeated sequence.

Repetition spacing in baboons

Experimental evidence about the effect of repetition spacing with nonverbal tasks and/or in nonhuman animals is very scarce. However, Ordonez Magro et al. (2022) recently conducted a study in Guinea baboons using a combination of the SRT task described in Section 2.2. and the Hebb learning paradigm (Hebb, 1961; Page et al., 2013) to understand how the memory trace of a repeated triplet consolidates with repetition but also, how spacing affects the trace development dynamics. Here, baboons still have to touch a red target on a touch screen which appears sequentially and have to process sequences of three touches (see Figure 1). On each trial, the sequences can be of two types: either a repeated triplet of touches (i.e., the Hebb sequence) or a triplet of random touches (i.e., filler sequences composed of positions that are different from the positions defining the Hebb sequence). The learning of the repeated triplets is assessed in baboons under four conditions: between two repetitions of the repeated sequence, there is either one filler sequence (1F-Condition), three filler sequences (3F-Condition), six filler sequences (6F-Condition), or no filler sequence (0F-Condition). As for Pinto Arata, Ramisch et al. (2024), a first result of this experiment is the replication of the predictability effect as evidenced by the decrease in their RTs for the second item and an even more pronounced decrease for the third item. The second main result concerns the repetition effect, with RTs for the unpredictable first item of the sequence also decreasing with repetition.

Furthermore, regarding the speed at which the memory trace is formed, the results show that the learning slopes for all items decrease as the number of interspersed filler sequences increases, indicating that learning is stronger when spacing is shorter. Finally, the most interesting feature of this study is that, by measuring the learning of the triplet

under increasing spacing conditions, authors are able to estimate the maximum spacing at which learning should still occur. They conclude that learning of the triplet can occur with spacing up to approximately 18 filler sequences, but is very unlikely to occur beyond that point. This rather precise prediction not only informs us about the survival of a transitory trace in memory, but can also be compared with predictions made by computational models.

Chunk size

Chunk size in the domain of language

Despite the fact that chunking helps to overcome the inherent limitations of memory storage and short-term working memory processing, it seems that the storage capacity of individual chunks themselves appears to be limited, with an average size of approximately 3-4 items (Allen & Coyne, 1988; Chase & Simon, 1973). Indeed, studies on deliberate chunking have shown that individuals tend to spontaneously form 3-item chunks (Johnson, 1970), and exhibit improved recall for clusters consisting of three to four items (e.g., Farrell, 2008; Parmentier & Maybery, 2008; Wickelgren, 1967).

In the linguistic domain, chunk size has also been studied mainly from the perspective of deliberate chunking, using tasks such as serial recall and immediate recall, especially with lists of words. For instance, Smalle et al. (2016) presented participants with either a list of isolated syllables or with a list where syllables were grouped into pairs by inserting a 1000 ms interstimulus interval every two syllables (i.e., an interstimulus interval was inserted after the 2nd, 4th, 6th, and 8th syllable). They found that participants showed better recall and faster learning for small chunks compared to larger chunks. In the case of automatic chunking, however, no experimental study to date has investigated the average

chunk size and the associated chunking dynamics and segmentation constraints during sequence learning.

Nevertheless, evidence from corpus analyses has shed new light on the question of chunk size in the case of natural language. For instance, Bannard and Matthews (2008) showed that there are many multiword sequences that occur almost as frequently or more frequently than single words in English, particularly multiword sequences of up to 4 words, suggesting that these sequences may be encoded in memory just like single words. Similarly, an analysis of the ORFÉO database (Benzitoun et al., 2016) - a French conversation corpus of 3 million words - carried out for this review, reveals the existence of relatively frequent multiword sequences composed of up to 9 words, but also shows that several multiword sequences longer than four words are much more frequent than smaller ones (see Figure 2). It is worth noting that the frequency of multiword sequences, like that of words, also appears to follow a Zipfian distribution (Ryland Williams et al., 2015). Thus, there are very few high-frequency multiword sequences that account for most of the utterances in conversation (e.g., in French: “parce que”, “Il y a”, “je ne sais pas”; in English: “such as”, “a lot of”, “I would like to”), and many low-frequency multiword sequences (e.g., in French: “comme beaucoup”, “pour ce qui”, “je ne veux pas”; in English: “make out”, “how much more”, “ I had no idea”). This type of analysis suggests that for relatively long multi-word sequences of very high frequency, the language processing system could develop representations of these sequences in the same way as word units of equivalent frequency. However, it remains to be shown that this hypothesis is empirically valid because, as we will see in the next section, the data we have on the

maximum size of a chunk in the field of visual-motor sequence learning indicates that, on average, there seems to be an intrinsic limit to the maximum size of a chunk.

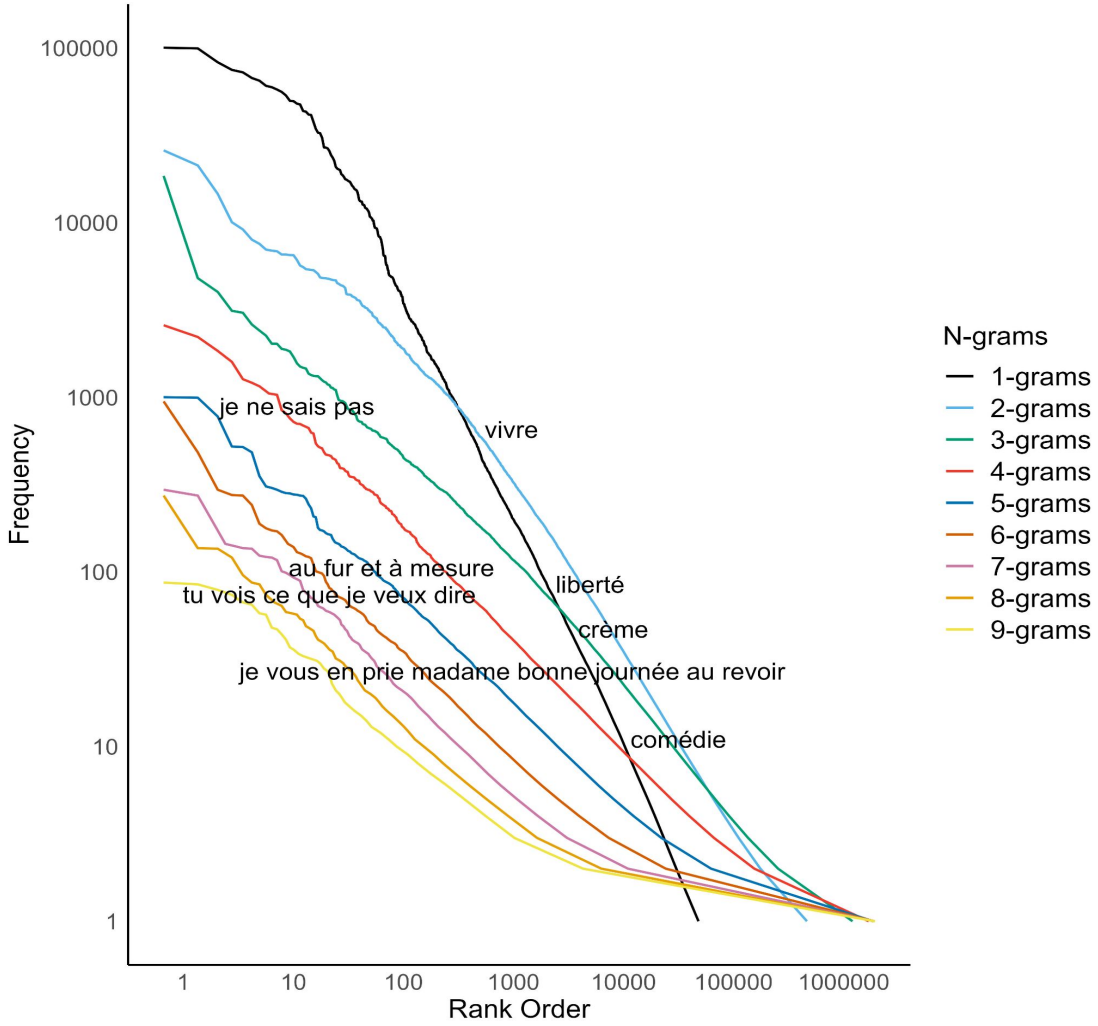


Figure 2. Frequency of n-grams plotted against their rank order, on logarithmic scales. Each line shows results for n-grams of a different length (one through nine words). Examples of multiword sequences of different length and single words are also plotted. The data were derived from the ORFÉO database.

Chunk size is generally limited and depends on several factors

In contrast to studies in the linguistic domain, the question of chunk size has already been investigated in visuo-motor sequence learning tasks. Evidence from this literature converges on a chunk size of about 3 to 4 items when testing humans (e.g., Abrahamse et

al., 2013; Verwey, 2001), but also non-human primates (Scarf et al., 2018). Some studies, however, sometimes report larger chunk sizes (e.g., Kennerley et al., 2004 observe chunks with a mean size of 8 items). While it has been shown that age differences can influence chunk size (Allen & Coyne, 1988; Allen & Crozier, 1992; Verwey et al., 2011), we argue here that chunk size tends to increase with practice, but there also seems to be an average limit to the size chunks can reach.

A first crucial factor to consider when measuring chunk size is the amount of expertise of the individual. Studying how expertise influences chunk size can be achieved by studying experts in their field, as in Gobet and Clarkson (2004), but also by measuring performance throughout extended practice of a sequence. Thus, a number of studies show that the chunking pattern evolves with practice and that chunks can be grouped together and become larger and fewer with a large number of repetitions (Acuna et al., 2014; Bera et al., 2021; Ramkumar et al., 2016; Tosatto, Fagot, et al., 2022; Tosatto, Fagot, Nemeth, et al., in press; Tosatto, Fagot, & Rey, 2023b, 2023a; Wymbs et al., 2012). Tosatto et al. (2022, 2023b) recently replicated these results in a comparative approach conducted on both humans and Guinea baboons using the SRT paradigm from Minier et al. (2016), where the same 9-item sequence was repeatedly presented for 1,000 trials to their participants. The results showed that the chunks became larger and fewer over repetitions, but also that the average size of the chunks was between 2 and 3 for humans, and 3 and 4 for baboons, even after a very large number of repetitions. The difference obtained between humans and non-human primates seems to be related to the speed-accuracy trade-off, which is not the same depending on the species, with humans achieving very fast execution speeds compared to monkeys (probably thanks to their language skills, which enable them to understand the structure of the experiment and adjust their performance accordingly). In fact, unlike monkeys, humans are likely to quickly realize that they are observing the same

repeated pattern throughout the task, leading to a change in the way the information is processed and the task is done. Once humans have noticed the repeated pattern, they will likely spend less time looking for the target on the screen and more time memorizing the execution of the motor sequence. This explicit awareness does not seem to occur in baboons, whose average execution times decrease more slowly and linearly (see Tosatto, Fagot, & Rey, 2023b, Figure 2), as if the visual target search stage were still unavoidable. In humans, explicit detection of the repetition seems to eliminate the visual search for the target in favor of an automated execution of the motor sequence, which is learned by heart. However, a potential consequence of this trade-off is that humans seem to create shorter chunks as they cannot respond any faster (i.e., floor effect) due to biomechanical constraints inherent in the task itself. Nevertheless, these data suggest that for these two species, the average size of the chunks reaches limits that cannot be exceeded, or only in exceptional cases.

Note that these studies also observed that the growth of chunks is governed by two reorganization mechanisms, observed in both species. Indeed, they found that chunks became larger via *concatenations* (i.e., two distinct chunks being grouped into one, Abrahamse et al., 2013; Wymbs et al., 2012) but may also evolve over repetitions through *recombinations* (i.e., a new segmentation pattern emerging, such as two chunks of 3 items becoming a chunk of 4 items and a chunk of 2 items). These identical dynamics being shared by the two species reinforces the idea that these reorganizations rely on the same elementary associative learning mechanisms.

Secondly, chunks do not only grow with practice, but it seems that their size limit is not dependent on sequence size. Tosatto, Fagot, Nemeth, et al. (in press) recently expanded on the study cited above that tested baboons on a 9-item sequence (i.e., Tosatto, Fagot, et al., 2022) and observe the evolution of chunks size during the extended learning of 4-item

or 5-item sequences. The results show that, even though chunks can hold 3 to 4 items in the case of long sequences of 9-items, shorter sequences of 4 or 5 items lead to similar size limits, even if the dynamics to reach this limit were not the same. These results therefore confirm that there is, on average, a limit to the size that chunks can reach, revealing a fundamental limit to chunk formation that should inform us about chunking mechanisms and their computational modeling.

Computational models of chunking

The three empirical phenomena described in Sections 2-4, i.e., predictability effects, the effect of repetition spacing, and the evolution of chunk size and chunk size limits, provide empirical data to constrain our understanding and modeling of chunking mechanisms. We will now see what current models of statistical learning, and more specifically, computational models of chunking, tell us about these three phenomena.

The theoretical proposals put forward to date attempt to explain how, within the flow of information we receive and process, the repetition of certain information is encoded by the brain and influences our subsequent performance on this information. Among all the proposals made, Swingley (2005) suggested distinguishing between bracketing and clustering approaches. Bracketing approaches assume that we learn the boundaries between the sequences we learn by repetition. The Simple Recurrent Network model (i.e., SRN, Elman, 1990) provides a computational implementation of this approach, using a relatively simple neural network that learns to predict outputs from inputs. When this type of network fails to clearly predict an output, we can then assume that we are at a boundary and that we need to segment processing at this point. Clustering approaches assume that we group the information that corresponds to a repeated sequence into a unit, also known as a chunk. The PARSER model (Perruchet & Vinter, 1998) is often cited as one of the models that implements this type of approach but today there exists several

variants of chunking models (e.g., Frank et al., 2010; French et al., 2011; Jessop et al., 2023; McCauley & Christiansen, 2019a; Monaghan & Christiansen, 2010).

The interest of such computational models is to eventually find situations where these models make orthogonal predictions which, in principle, should make it possible to falsify either approach. For example, Giroux and Rey (2009) noted that learning an ABC sequence leads the PARSER model to extract the ABC chunk and later process the sequence at chunk level, thus neglecting the processing of smaller units such as AB or BC. Conversely, SRN predicted that while learning ABC, this model continues to predict B well when A is presented as input and to predict C well when B is input. Results from an experiment with artificial languages showed instead that PARSER was right and that the smaller units that are included in the ABC chunk are processed less well over time. However, Plaut and Vande Velde (2017) demonstrated that testing the SRN model by taking into account whether the AB subunit is followed by a silence or the BC subunit is preceded by a silence, allowed Giroux and Rey's (2009) data to be accounted for, thus negating PARSER's advantage.

Converging evidence has shown that chunk-based models provide a better fit to human sequence learning data than alternative models in the statistical learning literature, such as those based on statistical computations using transitional probabilities (Emerson & Conway, 2023; Isbilen et al., 2020; Perruchet, 2019; Perruchet et al., 2014). The aim of this section is to examine the extent to which chunk-based models can simulate the findings presented in this review.

Predictability effect

Concerning the predictability effect (i.e., the faster decrease in response times on C than on B during repeated exposure to the ABC triplet), Minier et al. (2016) suggested that SRN better accounts for this result due to the existence of a layer of contextual units within this

neural network. These units contain the copy of the activity of the hidden units at time *t-1* and thus make it possible to take into account more distant contextual information. When processing an ABC triplet, when B is an input of the network and C is the expected output, context units will take into account the activity of the hidden layer at time *t-1*, i.e., when A was an input and B was the expected output. Therefore, the output C is predicted both by input activity produced by B and by the additional contextual information coming from A which is coded by the layer of contextual units. SRN thus seems to capture predictability effects in the processing of repeated ABC triplets.

In this situation, it is hard to know what PARSER predicts, as it is relatively difficult to interrogate this model in a task where learning is measured online by studying the evolution of response times at each position of the triplet. Assuming that PARSER extracts chunks of different sizes over time and that these chunks help predict the next elements of a sequence, we can assume that a chunk corresponding to the target triplet ABC will be quickly extracted. However, there's nothing in PARSER to indicate whether differential processing should be assumed for items B and C of the ABC triplet. So it's not clear whether PARSER can account for this predictability effect.

On the other hand, a chunking model such as TRACX (French et al., 2011) is much clearer in its predictions, since it assumes a left-to-right learning mechanism where AB will systematically be learned before BC, thus predicting faster learning on B than C, which unfortunately is at odds with the experimental data and thus seems to falsify a central assumption of this model. Similarly, chunking models from the EPAM/CHREST family, such as MOSAIC (Freudenthal et al., 2007, 2009, 2023) and CLASSIC (Cabiddu et al., 2023), process sequences from left to right but predict faster learning on C because they learn from the right edge of the sequence. However, these models have mainly been used to account for syntax acquisition and early word segmentation. It is unclear how they would

behave with other types of stimuli, hence further research is needed to test these models in other learning situations.

More generally, like PARSER, all the models that assume the creation of chunks as traces in long-term memories do not seem to explain the predictability effect easily, because they do not say anything very precise about the processing of the elements that make up the chunk. Connectionist models of chunking such as TRACX could give more precise information, but unfortunately, they predict the opposite effect to what is observed in the data.

Recent theoretical proposals based on the principles of Hebbian learning could provide more precise explanations of chunking mechanisms, which can be thought of as resulting from elementary associative mechanisms (Brunel & Lavigne, 2009; Endress & Johnson, 2021; Köksal Ersöz et al., 2020; Lavigne et al., 2014; Tovar et al., 2018; Tovar & Westermann, 2017, 2023). In these models, each element of a sequence is encoded by a population of neurons, and connections between populations of neurons evolve according to the principles of Hebbian learning. Basically, if two populations are frequently co-active, their connections strengthen, otherwise they weaken. If we assume that such a Hebbian network allows us to associate the population coding for A with that coding for B, but also that coding for B with that coding for C, we cannot account for the predictability effect because connections between populations of neurons will increase with the same strength. On the other hand, if we assume that the population coding for A still remains sufficiently active when the population coding for C is active, then we can imagine Hebbian learning between the A and C populations. In this way, such a network would

predict an advantage of C over B due to a greater amount of contextual information predicting C than B.

Furthermore, Hebbian networks could also be used to model different experimental factors that may influence the predictability effect, such as the elapsed time between the presentation of the repeated sequence. For instance, the elapsed time between the presentation of A and C was shorter in Rey et al.'s (2020) task than in Pinto Arata, Ordonez Magro et al.'s (in press) task. In fact, the repeated triplets in Pinto Arata, Ordonez Magro et al. (in press) consisted of words rather than letters, which may have resulted in participants spending more time processing each element of the triplet, leading to a lower co-activation between A and C. Hebbian networks, such as Tovar and Westermann's (2023) model, can easily model these different results by using different activation decay parameters to capture the elapsed time between stimuli. Clearly, unlike chunking models such as PARSER, this type of formalism could provide a better description of the fine dynamics of these associative chunking mechanisms.

Spacing effect

The spacing effect is an excellent way of studying the development of a chunk's memory trace in long-term memory. The experiments described in Section 3 help to constrain the way in which chunking models account for the progressive development of a chunk as the information that characterizes it is repeated. In such studies, where, for example, an ABC sequence is repeated a certain number of times, it is possible to study the dynamics of the simple repetition effect by tracking the evolution of response times (RTs) on position A of the triplet whose occurrence is not predictable. Similarly, by

studying RTs on predictable positions of the triplet (i.e., B and C), it is possible to characterize precisely how the repetition effect interacts with the predictability effect.

What's clear from these empirical studies is that the smaller the spacing, the stronger the chunk's memory trace. This suggests that the reinforcement of a memory trace depends on the frequency with which it is repeated. If the spacing is too great, the trace benefits less from repetition, probably because between two repetitions, the chunk trace has lost its strength and undergone a form of erosion or forgetting. Chunking models like PARSER assume forgetting processes by decay or interference. For example, decay is implemented by a linear decrease in the strength of a chunk unit, which is probably implausible as chunks are quickly forgotten in this way. Similarly, models such as PUDDLE (Monaghan & Christiansen, 2010) and CIPAL (Jessop et al., 2023) assume no forgetting and state that when a chunk is created, its trace in memory can only increase and never decrease or disappear. Here too, this type of assumption seems incompatible with the effects of repetition spacing, which seem to indicate that the increase in the strength of a chunk's trace is linked to the spacing between two repetitions of the chunk. Experiments on repetition spacing are therefore crucial to help us parameterize and characterize forgetting in computational models.

Chunk size

The data we have to date suggests that there is an intrinsic limit to the size of a chunk. Numerous studies agree that automatic chunking can group an average of 3 to 4 elements within a chunk (e.g., Tosatto, Fagot, Nemeth, et al., in press). However, no chunking model sets such a limit. For example, PARSER in no way limits the possibility of concatenating multiple long elements already stored in long-term memory, and this model even allows to create chunks of chunks of chunks despite their size. Statistical learning models such as SRN could perhaps account for such a limit, as the information copied into the

contextual layers of the network does not a priori allow infinite consideration of previous contextual information.

On the other hand, recent statistical learning models based on the principles of Hebbian learning could perhaps provide the beginnings of an answer to this fundamental problem (e.g., Tovar & Westermann, 2023). Indeed, assuming that each element of a five-item sequence ABCDE is encoded by a specific population of neurons, we can foresee that the population encoding A will probably no longer be active when the population encoding E becomes active in turn. In this way, populations A and E might never be associated, as the principles of Hebbian learning, which presuppose coactivation of these two populations, would not apply in this scenario. This would provide an explanation for the intrinsic limitation that chunks seem to have.

Discussion

Statistical learning has been proposed to rely on general memory processes and associative learning mechanisms, particularly chunking. In the current paper, we examined three empirical phenomena present during sequence learning, namely predictability effect, repetition spacing and chunk size (see Figure 3 for an overview of the studies reviewed). The goal of this review was to provide a thorough characterization of chunking as a domain-general learning mechanism and to shed light on its associative properties. To this end, we compared humans to non-human primates' performances across different experimental tasks. Furthermore, we discussed the extent to which chunk-based models of statistical learning account for these three phenomena.

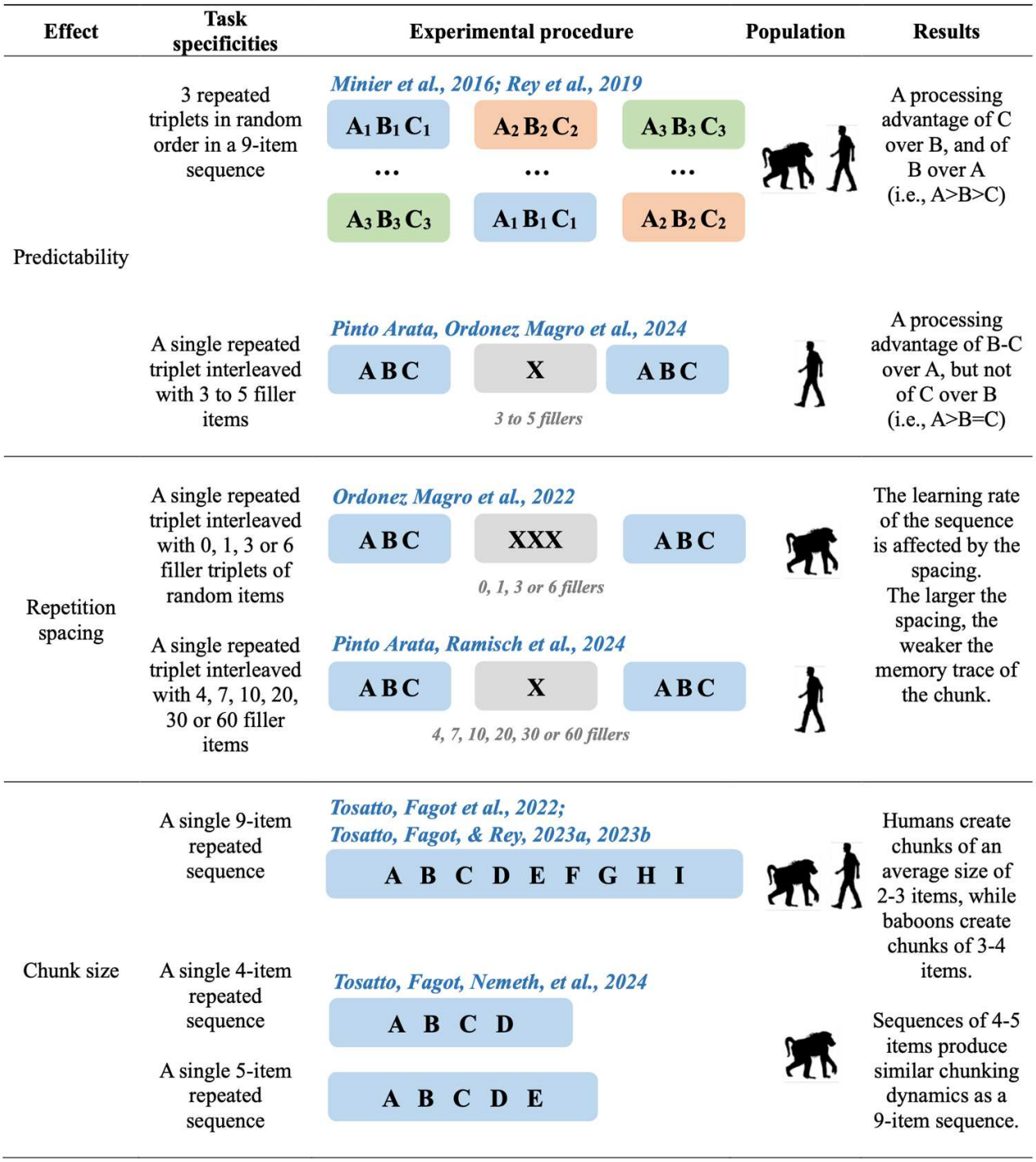


Figure 3. Summary of experiments reviewed. The table shows each effect reviewed along with tasks specificities, the experimental procedure used, and the population tested. Each experimental procedure is presented with the corresponding paper.

In the past two decades, research has consistently shown that chunking can successfully account for sequence learning across a wide range of domains and stimuli.

Nevertheless, our results somewhat challenge the explanatory power of this mechanism, as certain learning dynamics and chunk reorganization mechanisms cannot be simulated by existing chunk-based models. Similarly, chunking has been found to exhibit limitations in more complex learning situations, such as the learning of non-adjacent dependencies (e.g., Bonatti et al., 2006; Endress & Bonatti, 2007). In fact, most research on chunking has primarily focused on the processing of adjacent regularities, where co-occurring elements appear systematically one after the other in the environment (e.g., A and B in the sequence AB). However, real-world regularities can also exhibit non-adjacent patterns, where elements do not directly follow each other in a sequence (e.g., the sequence AXB, where A and B always co-occur, but X varies). Crucially, non-adjacent dependencies are pervasive in the environment and particularly prominent in natural language. For instance, in English, the auxiliary and the inflectional morpheme co-occur at a distance, whereas the verbal root can vary (e.g., "*is _ing*" => *is walking*, *is looking*). The detection and extraction of non-adjacent dependencies is thought to be essential for making structural generalizations across regularities (Frost et al., 2019; Grama & Wijnen, 2018).

Despite numerous studies showing that adults (e.g., Frost & Monaghan, 2016; Isbilen et al., 2022; Perruchet et al., 2004; Romberg & Saffran, 2013; Vilkaitė, 2016), infants (e.g., Frost et al., 2020; Marchetto & Bonatti, 2015; Van Der Kant et al., 2020), and non-human primates can successfully acquire non-adjacent dependencies (e.g., Malassis et al., 2018; Sonnweber et al., 2015; Watson et al., 2020), learning this type of regularities has proven to be very challenging and dependent on a number of facilitating factors (Gebhart et al., 2009; Newport & Aslin, 2004; Tosatto, Bonafos, et al., 2022; Weyers et al., 2022; see Wilson et al., 2020 for a review). Some authors have highlighted the inability of chunk-based models to learn non-adjacent dependencies, as chunking relies on the recognition of previously encountered sequences. Indeed, the presence of variable elements between

co-occurring elements would hinder chunk formation (Bonatti et al., 2006; Kuhn & Dienes, 2005). Perhaps, the only chunk-based model capable of learning this type of regularities is the chunk-based learner (CLB) model developed by McCauley and Christiansen (2019b), which has been shown to be able to track and generalize over non-adjacent dependencies. However, despite its name, it is a hybrid model in which initial chunk formation relies on both chunking and the statistical computation of pairwise associations (i.e., backward transitional probabilities). Consequently, the extent to which non-adjacent dependencies can be acquired solely through chunking remains a matter of debate (Endress & Bonatti, 2016).

However, as shown by Pacton and Perruchet (2008, see also Pacton et al., 2015), one way of reintroducing the idea of chunking in the case of non-adjacent dependencies comes from the idea that even far away, certain information can be processed adjacently by paying attention to it almost jointly. This brings us back to the Hebbian idea of assuming that two pieces of information, encoded by two populations of neurons, will be associated if they are coactive at the same time. So, if we draw a parallel between the idea put forward by Pacton and Perruchet (2008) of simultaneous attentional processing of non-adjacent information, which enables this information to be chunked, and the Hebbian idea of the coactivation of this information leading to a strengthening of their connections, we find an explanation in terms of the chunking of non-adjacent dependencies.

The empirical data on visuomotor learning that we have reviewed so far suggest that the size of chunks seems to be limited to 3-4 items. However, generalizing this limit to language is not as straightforward as it might seem. In fact, defining the items that constitute a chunk can prove difficult and raises a number of issues. For starters, if by "item" we denote a smaller chunk or sub-chunk in a sequence (e.g., A in an ABC chunk), then what should be considered an item in a linguistic sequence becomes fuzzy. To

illustrate this issue, let us take the multiword phrase “We haven't got enough” as an example. When this phrase is analyzed, it can be broken down into four words, into six syllables, or even into the 14 phonemes that make up the phrase, making it difficult to know what is really represented in memory. Second, if chunking is viewed as an incremental process in which items are learned and progressively chunked into larger sequences, such as in PARSER, the question of chunk size becomes increasingly difficult as chunks keep evolving over time. This is especially the case since many chunking models assume that chunks form a hierarchy of different-sized chunks that interact during learning (e.g., CLASSIC: Cabiddu et al., 2023; PARSER: Perruchet & Vinter, 1998; MDLChunker: Robinet et al., 2011).

Therefore, a goal of future research will be to investigate which items are used to build chunk hierarchies, how chunk size is determined by our cognitive system, and whether there is a limit to chunk size. Norris and Kalm (2021) have suggested that one possible way to determine memory capacity, and hence chunk size, is to measure information in “bits” and compressibility. In this account, memory capacity results from the interaction of the nature of the information (e.g., phonemes, words) encoded in bits and the efficiency of data compression. Thus, chunk size depends on how effectively the information in memory can be compressed. It is worth noting that there seems to exist a bias toward creating short rather than long chunks, even though it has been shown that learners can create fairly long chunks (e.g., Kennerley et al., 2004). For instance, Frank et al. (2010) presented participants with an artificial language that could be segmented into disyllabic, trisyllabic, or hexasyllabic chunks. They found that participants preferred disyllabic and trisyllabic segmentations over hexasyllabic ones. This is likely due to the participants'

prior linguistic experience, which may shape segmentation patterns, but also to the fact that short chunks may help maximize learnability.

These results on language segmentation, together with the fact that our corpus analysis seems to reveal the existence of relatively frequent long multiword sequences, lead to questions about how the nature of the linguistic information being processed affects chunking, for instance: What are the features that can lead to the formation of long rather than short multiword sequences? Are these features the same across languages? Or are all multiword sequences composed of relatively short words? If not, what are the factors that modulate their composition? We believe that addressing these issues is crucial to understanding how our cognitive system processes information in real time, but cannot be done without developing models that take into account both the psychological and biological bases of learning.

From a theoretical perspective, the concept of chunking principally refers to the psychological process of grouping information in order to bypass short-term memory limitations and ultimately form high-level representations of that information in long-term memory. Even though chunking is considered to rely on associative principles, no biological account to date has yet been proposed to explain exactly how this process occurs at the neural level. We have seen that neurocomputational models grounded in Hebbian learning principles may bridge this gap and offer a promising alternative to chunking. Such models have the potential to provide a more plausible characterization of statistical learning mechanisms and shed light on the dynamic formation of memory traces in the brain from a biological perspective (e.g., Tovar & Westermann, 2023). Note that the proposals we have made to account for the predictability effect or the question of the intrinsic limit of chunk size within this theoretical framework are open to debate, and that other computational solutions are undoubtedly possible. For example, to account for

the predictability effect, we can also assume that the adjacent presentation of AB leads to the emergence of a new population of neurons that specifically encode AB, which establish a synaptic connection with C, producing an AB-C association (Bourjaily & Miller, 2011; Lavigne et al., 2014; Miyashita, 2004). Similarly, reorganization mechanisms could be explained by the interaction of neighboring populations of neurons that encode different chunks. This interaction induces a modification of their connection weights, leading to the development of new synapses between them, resulting in chunk rearrangements.

These neurocomputational models are also well suited for modeling repetition spacing effects due to the possibility to modulate learning rate during sequence learning by simulating long-term potentiation (LTP) and long-term depression (LTD), two cellular mechanisms that strengthen or weaken the synaptic connection between populations of neurons, through co-activation thresholds and synaptic weight decay parameters (e.g., Tovar et al., 2018; Tovar & Westermann, 2023).

Conclusion

In this comprehensive review, we have presented compelling evidence that predictability effect, repetition spacing, and chunk size exert similar influences on chunking in humans and non-human primates during sequence learning. These findings provide further support for the existence of shared associative learning mechanisms between the two species and shed new light on the associative properties of chunking. However, we have also shown that chunk-based models are unable to fully explain these phenomena. Therefore, we argue for a shift towards neurocomputational models based on Hebbian learning principles, which may provide a more comprehensive understanding of the underlying associative processes involved in statistical learning.

We believe that the main challenge for future research is to elucidate how these models can effectively account not only for the phenomena discussed in this paper, but also for

other key findings in the statistical learning literature. For instance, the relationship between forward transitional probabilities and frequency (e.g., Onnis & Huettig, 2021), the extraction of backward transitional probabilities (e.g., Chartier & Dautriche, 2023; McCauley & Christiansen, 2019a; Pelucchi et al., 2009; Perruchet & Desaulty, 2008), the extraction of high-order transitional probabilities (e.g., Lazartigues et al., 2021, 2022; Rey et al., 2022), and the development of stimulus equivalence in humans and animals (e.g., Chartier & Fagot, 2022a, 2022b). Together, these empirical data and new computational approaches should enable us to better understand the essential associative mechanisms of chunking that structure our mental life, and more specifically the implicit statistical learning of sequences.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

LPA was supported by a doctoral fellowship of the French Ministry of Higher Education, Research, and Innovation. This research was supported by the Convergence Institute ILCB (ANR-16-CONV-0002), the centre for research in education Ampiric, the COMPO ANR project (#ANR-23-CE23-0031) and the HEBBIAN ANR project (#ANR-23-CE28-0008).

References

Abrahamse, E. L., Ruitenberg, M. F. L., de Kleine, E., & Verwey, W. B. (2013). Control of automated behavior: Insights from the discrete sequence production task. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00082

Acuna, D. E., Wymbs, N. F., Reynolds, C. A., Picard, N., Turner, R. S., Strick, P. L., Grafton, S. T., & Kording, K. P. (2014). Multifaceted aspects of chunking enable robust algorithms. *Journal of Neurophysiology*, *112*(8), 1849–1856. https://doi.org/10.1152/jn.00028.2014

Allen, P. A., & Coyne, A. C. (1988). Age differences in primary organization or processing variability? Part I: An examination of age and primary organization. *Experimental Aging Research*, *14*(3), 143–149. https://doi.org/10.1080/03610738808259739

Allen, P. A., & Crozier, L. C. (1992). Age and Ideal Chunk Size. *Journal of Gerontology*, *47*(1), P47–P51. https://doi.org/10.1093/geronj/47.1.P47

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth—Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107–129. https://doi.org/10.1080/15475441.2010.505489

Arnon, I., & Priva, U. C. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*(3), 349–371. https://doi.org/10.1177/0023830913484891

Avarguès-Weber, A., Finke, V., Nagy, M., Szabó, T., D'Amaro, D., Dyer, A. G., & Fiser, J. (2020). Different mechanisms underlie implicit visual statistical learning in honey bees and humans. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/PNAS.1919387117

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning of four-word combinations. *Psychological Science*, *19*(3), 241–248. https://doi.org/10.1111/j.1467-9280.2008.02075.x

Basyouni, R., & Parkinson, C. (2022). Mapping the social landscape: Tracking patterns of interpersonal relationships. *Trends in Cognitive Sciences*, *26*(3), 204–221. https://doi.org/10.1016/j.tics.2021.12.006

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. *Topics in Cognitive Science*, *11*(3), 482–503. https://doi.org/10.1111/tops.12420

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2017). Implicit and explicit contributions to statistical learning. *Physiology & Behavior*, *176*(5), 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

Benzitoun, C., Debaisieux, J.-M., & Deulofeu, H.-J. (2016). Le projet ORFÉO: Un corpus d'étude pour le français contemporain. *Corpus*, *15*. https://doi.org/10.4000/corpus.2936

Bera, K., Shukla, A., & Bapi, R. S. (2021). Motor chunking in internally guided sequencing. *Brain Sciences*, *11*(3), 292. https://doi.org/10.3390/brainsci11030292

Bonatti, L. L., Nespor, M., Peña, M., & Mehler, J. (2006). How to hit scylla without avoiding charybdis: Comment on Perruchet, Tyler, Galland, and Peereman (2004). *Journal of Experimental Psychology: General*, *135*(2), 314–321. https://doi.org/10.1037/0096-3445.135.2.314

Bourjaily, M. A., & Miller, P. (2011). Synaptic plasticity and connectivity requirements to produce stimulus-pair specific responses in recurrent networks of spiking neurons. *PLoS Computational Biology*, *7*(2), e1001091. https://doi.org/10.1371/journal.pcbi.1001091

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind: Discoveries in Cognitive Science*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026

Brand, M., Giroux, I., Puijalon, C., & Rey, A. (2007). Syllable onsets are perceptual reading units. *Memory & Cognition*, *35*(5), 966–973. https://doi.org/10.3758/BF03193469

Brunel, N., & Lavigne, F. (2009). Semantic priming in a cortical network model. *Journal of Cognitive Neuroscience*, *21*(12), 2300–2319. https://doi.org/10.1162/jocn.2008.21156

Cabiddu, F., Bott, L., Jones, G., & Gambi, C. (2023). Classic utterance boundary: A chunking-based model of early naturalistic word segmentation. *Language Learning*, 1–34. https://doi.org/10.1111/lang.12559

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, *63*(1), 95–122. https://doi.org/10.1177/0023830918823230

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Chartier, T. F., & Dautriche, I. (2023). Do backward associations have anything to say about language? *Cognitive Science*, *47*(4), e13282. https://doi.org/10.1111/cogs.13282

Chartier, T. F., & Fagot, J. (2022a). Associative symmetry: A divide between humans and nonhumans? *Trends in Cognitive Sciences*, *26*(4), 286–289. https://doi.org/10.1016/j.tics.2022.01.009

Chartier, T. F., & Fagot, J. (2022b). Simultaneous learning of directional and non-directional stimulus relations in baboons (Papio papio). *Learning & Behavior*. https://doi.org/10.3758/s13420-022-00522-8

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*(3), 468–481. https://doi.org/10.1111/tops.12332

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, 1–72. https://doi.org/10.1017/S0140525X1500031X

Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience and Biobehavioral Reviews*, *112*, 279–299. https://doi.org/10.1016/j.neubiorev.2020.01.032

Conway, C. M., Eghbalzad, L., Deocampo, J. A., Smith, G. N. L., Na, S., & King, T. Z. (2020). Distinct neural networks for detecting violations of adjacent versus nonadjacent sequential dependencies: An fMRI study. *Neurobiology of Learning and Memory*, *169*, 107175. https://doi.org/10.1016/j.nlm.2020.107175

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Daltrozzo, J., Emerson, S. N., Deocampo, J., Singh, S., Freggens, M., Branum-Martin, L., & Conway, C. M. (2017). Visual statistical learning is related to natural language ability in adults: An ERP study. *Brain and Language*, *166*, 40–51. https://doi.org/10.1016/j.bandl.2016.12.005

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Emerson, S. N., & Conway, C. M. (2023). Chunking versus transitional probabilities: Differentiating between theories of statistical learning. *Cognitive Science*, *47*(5), e13284. https://doi.org/10.1111/cogs.13284

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247–299. https://doi.org/10.1016/j.cognition.2006.09.010

Endress, A. D., & Bonatti, L. L. (2016). Words, rules, and mechanisms of language acquisition. *WIREs Cognitive Science*, *7*(1), 19–35. https://doi.org/10.1002/wcs.1376

Endress, A. D., & Johnson, S. P. (2021). When forgetting fosters learning: A neural network model for statistical learning. *Cognition*, *213*. https://doi.org/10.1016/j.cognition.2021.104621

Farrell, S. (2008). Multiple roles for time in short-term memory: Evidence from serial recall of order and timing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 128–145. https://doi.org/10.1037/0278-7393.34.1.128

Fonollosa, J., Neftci, E., & Rabinovich, M. (2015). Learning of chunking sequences in cognition and behavior. *PLOS Computational Biology*, *11*(11), e1004592. https://doi.org/10.1371/journal.pcbi.1004592

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107–125. https://doi.org/10.1016/j.cognition.2010.07.005

Frank, M. C., Tily, H., Arnon, I., & Goldwater, S. (2010). Beyond transitional probabilities: Human learners impose a parsimony bias in statistical word segmentation. *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614–636. https://doi.org/10.1037/a0025255

Freudenthal, D., Gobet, F., & Pine, J. M. (2023). MOSAIC+: A crosslinguistic model of verb-marking errors in typically developing children and children with developmental language disorder. *Language Learning*, 1–35. https://doi.org/10 . 1111 / l a n g . 12580

Freudenthal, D., Pine, J. M., Aguado-Orea, J., & Gobet, F. (2007). Modeling the developmental patterning of finiteness marking in English, Dutch, German, and Spanish using MOSAIC. *Cognitive Science*, *31*(2), 311–341. https://doi.org/10.1080/15326900701221454

Freudenthal, D., Pine, J. M., & Gobet, F. (2009). Simulating the referential properties of Dutch, German, and English root infinitives in MOSAIC. *Language Learning and Development*, *5*(1), 1–29. https://doi.org/10.1080/15475440802502437

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Frost, R. L. A., Isbilen, E. S., Christiansen, M. H., & Monaghan, P. (2019). Testing the limits of non-adjacent dependency learning: Statistical segmentation and generalization across domains. *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, 1787–1793.

Frost, R. L. A., Jessop, A., Durrant, S., Peter, M. S., Bidgood, A., Pine, J. M., Rowland, C. F., & Monaghan, P. (2020). Non-adjacent dependency learning in infancy, and its link to language development. *Cognitive Psychology*, *120*, 101291. https://doi.org/10.1016/j.cogpsych.2020.101291

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70–74. https://doi.org/10.1016/j.cognition.2015.11.010

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, *16*(3), 486–490. https://doi.org/10.3758/PBR.16.3.486

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science*, *33*(2), 260–272. https://doi.org/10.1111/j.1551-6709.2009.01012.x

Gobet, F. (2017). Entrenchment, Gestalt formation, and chunking. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 245–267). American Psychological Association. https://doi.org/10.1037/15969-012

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four … or is it two? *Memory*, *12*(6), 732–747. https://doi.org/10.1080/09658210344000530

Gobet, F., Lane, P., Croker, S., Cheng, P., Jones, G., Oliver, I., & Pine, J. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*(6), 236–243. https://doi.org/10.1016/S1364-6613(00)01662-4

Gobet, F., Lloyd-Kelly, M., & Lane, P. C. R. (2016). What's in a name? The multiple meanings of "Chunk" and "Chunking." *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00102

Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory*, *6*(3), 225–255. https://doi.org/10.1080/741942359

Grama, I., & Wijnen, F. (2018). Learning and generalizing non-adjacent dependencies in 18-month-olds: A mechanism for language acquisition? *PLOS ONE*, *13*(10), e0204481. https://doi.org/10.1371/journal.pone.0204481

Havron, N., & Arnon, I. (2020). Starting big: The effect of unit size on language learning in children and adults. *Journal of Child Language*, *48*(2), 244–260. https://doi.org/10.1017/S0305000920000264

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning: A Symposium* (pp. 37–46).

Hu, Y., & Ericsson, K. A. (2012). Memorization and recall of very long lists accounted for within the Long-Term Working Memory framework. *Cognitive Psychology*, *64*(4), 235–266. https://doi.org/10.1016/j.cogpsych.2012.01.001

Isbilen, E. S., & Christiansen, M. H. (2018). Chunk-based memory constraints on the cultural evolution of language. *Topics in Cognitive Science*, *12*, 713–726. https://doi.org/10.1111/tops.12376

Isbilen, E. S., Frost, R. L. A., Monaghan, P., & Christiansen, M. H. (2022). Statistically based chunking of nonadjacent dependencies. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001207

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science*, *44*(7). https://doi.org/10.1111/cogs.12848

Jessop, A., Pine, J., & Gobet, F. (2023). *Chunk-based Incremental Processing and Learning: An integrated theory of word discovery, vocabulary growth, and speed of lexical processing*. PsyArXiv. https://doi.org/10.31234/osf.io/dukpt

Johnson, N. F. (1970). The role of chunking and organization in the process of recall. In *Psychology of Learning and Motivation* (Vol. 4, pp. 171–247). Academic Press. https://doi.org/10.1016/S0079-7421(08)60432-6

Kennerley, S. W., Sakai, K., & Rushworth, M. F. S. (2004). Organization of action sequences and the role of the pre-SMA. *Journal of Neurophysiology*, *91*(2), 978–993. https://doi.org/10.1152/jn.00651.2003

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, *461*(2), 145–149. https://doi.org/10.1016/j.neulet.2009.06.030

Köksal Ersöz, E., Aguilar, C., Chossat, P., Krupa, M., & Lavigne, F. (2020). Neuronal mechanisms for sequential activation of memory items: Dynamics and reliability. *PLOS ONE*, *15*(4), e0231165. https://doi.org/10.1371/journal.pone.0231165

Kuhn, G., & Dienes, Z. (2005). Implicit learning of nonlocal musical rules: Implicitly learning more than chunks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1417–1432. https://doi.org/10.1037/0278-7393.31.6.1417

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Lavigne, F., Avnaïm, F., & Dumercy, L. (2014). Inter-synaptic learning of combination rules in a cortical network model. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00842

Lazartigues, L., Mathy, F., & Lavigne, F. (2021). Statistical learning of unbalanced exclusive-or temporal sequences in humans. *PLOS ONE*, *16*(2), e0246826. https://doi.org/10.1371/journal.pone.0246826

Lazartigues, L., Mathy, F., & Lavigne, F. (2022). Probability, dependency, and frequency are not all equally involved in statistical learning. *Experimental Psychology*, *69*(5), 241–252. https://doi.org/10.1027/1618-3169/a000561

Malassis, R., Rey, A., & Fagot, J. (2018). Non-adjacent dependencies processing in human and non-human primates. *Cognitive Science*, *42*(5), 1677–1699. https://doi.org/10.1111/cogs.12617

Marchetto, E., & Bonatti, L. L. (2015). Finding words and word structure in artificial speech: The development of infants' sensitivity to morphosyntactic regularities. *Journal of Child Language*, *42*(4), 873–902. https://doi.org/10.1017/S0305000914000452

Mareschal, D., & French, R. M. (2017). TRACX2: A connectionist autoencoder using graded chunks to model infant visual statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160057. https://doi.org/10.1098/rstb.2016.0057

Mathy, F., Fartoukh, M., Gauvrit, N., & Guida, A. (2016). Developmental abilities to form chunks in immediate memory and its non-relationship to span development. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00201

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, *122*(3), 346–362. https://doi.org/10.1016/j.cognition.2011.11.003

McCauley, S. M., & Christiansen, M. H. (2019a). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, *126*(1), 1–51. https://doi.org/10.1037/rev0000126

McCauley, S. M., & Christiansen, M. H. (2019b). Modeling children's early linguistic productivity through the automatic discovery and use of lexically-based frames.

Proceedings of the 41st Annual Conference of the Cognitive Science Society, 782–788.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *2*, 1–21.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Minier, L., Fagot, J., & Rey, A. (2016). The temporal dynamics of regularity extraction in non-human primates. *Cognitive Science*, *40*(4), 1019–1030. https://doi.org/10.1111/cogs.12279

Miyashita, Y. (2004). Cognitive memory: Cellular and network machineries and their top-down control. *Science*, *306*(5695), 435–440. https://doi.org/10.1126/science.1101864

Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, *37*(3), 545–564. https://doi.org/10.1017/S0305000909990511

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Norris, D., & Kalm, K. (2021). Chunking and data compression in verbal short-term memory. *Cognition*, *208*. https://doi.org/10.1016/j.cognition.2020.104534

Onnis, L., & Huettig, F. (2021). Can prediction and retrodiction explain whether frequent multi-word phrases are accessed 'precompiled' from memory or compositionally constructed on the fly? *Brain Research*, *1772*. https://doi.org/10.1016/j.brainres.2021.147674

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, *105*(7), 2745–2750. https://doi.org/10.1073/pnas.0708424105

Ordonez Magro, L., Fagot, J., Grainger, J., & Rey, A. (2022, June 1–3). *Spacing of repetitions in implicit statistical learning* [Poster presentation]. Interdisciplinary Advances in Statistical Learning, San Sebastian, Spain.

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 80–96. https://doi.org/10.1037/0278-7393.34.1.80

Pacton, S., Sobaco, A., & Perruchet, P. (2015). Is an attention-based associative account of adjacent and nonadjacent dependency learning valid? *Acta Psychologica*, *157*, 195–199. https://doi.org/10.1016/j.actpsy.2015.03.002

Page, M. P. A., Cumming, N., Norris, D., McNeil, A. M., & Hitch, G. J. (2013). Repetition-spacing and item-overlap effects in the Hebb repetition task. *Journal of Memory and Language*, *69*(4), 506–526. https://doi.org/10.1016/j.jml.2013.07.001

Parmentier, F. B. R., & Maybery, M. T. (2008). Equivalent effects of grouping by time, voice, and location on response timing in verbal serial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1349–1355. https://doi.org/10.1037/a0013258

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244–247. https://doi.org/10.1016/j.cognition.2009.07.011

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Science*, *11*(3), 520–535. https://doi.org/10.1111/tops.12403

Perruchet, P., & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation ? *Memory and Cognition*, *36*(7), 1299–1305. https://doi.org/10.3758/MC.36.7.1299

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Perruchet, P., Poulin-Charronnat, B., Tillmann, B., & Peereman, R. (2014). New evidence for chunk-based models in word segmentation. *Acta Psychologica*, *149*, 1–8. https://doi.org/10.1016/j.actpsy.2014.01.015

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General*, *133*(4), 573–583. https://doi.org/10.1037/0096-3445.133.4.573

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246–263. https://doi.org/10.1006/jmla.1998.2576

Perruchet, P., & Vinter, A. (2002). The self-organizing consciousness. *Behavioral and Brain Sciences*, *25*(3), 297–330. https://doi.org/10.1017/S0140525X02000067

Perry, C. (2023). Graphemes are used when reading: Evidence from Monte Carlo simulation using word norms from mega-studies. *Quarterly Journal of*

Experimental Psychology, *76*(2), 419–428. https://doi.org/10.1177/17470218221086533

Pinto Arata, L., Ordonez Magro, L., Ramisch, C., Grainger, J., & Rey, A. (2024). The dynamics of multiword sequence extraction. *Quarterly Journal of Experimental Psychology*, 1–24. https://doi.org/10.1177/17470218241228548

Pinto Arata, L., Ramisch, C., & Rey, A. (2024, June 5–7). *The effect of repetition spacing in multiword sequence learning* [Poster presentation]. Interdisciplinary Advances in Statistical Learning, San Sebastian, Spain.

Plaut, D. C., & Vande Velde, A. K. (2017). Statistical learning of parts and wholes: A neural network approach. *Journal of Experimental Psychology: General*, *146*(3), 318–336. https://doi.org/10.1037/xge0000262

Pulido, M. F. (2021). Individual chunking ability predicts efficient or shallow L2 processing: Eye-tracking evidence from multiword units in relative clauses. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.607621

Ramkumar, P., Acuna, D. E., Berniker, M., Grafton, S. T., Turner, R. S., & Kording, K. P. (2016). Chunking as the result of an efficiency computation trade-off. *Nature Communications*, *7*(1), Article 1. https://doi.org/10.1038/ncomms12176

Rey, A., Bogaerts, L., Tosatto, L., Bonafos, G., Franco, A., & Favre, B. (2020). Detection of regularities in a random environment. *Quarterly Journal of Experimental Psychology*, *73*(12), 1–13. https://doi.org/10.1177/1747021820941356

Rey, A., Fagot, J., Mathy, F., Lazartigues, L., Tosatto, L., Bonafos, G., Freyermuth, J., & Lavigne, F. (2022). Learning higher-order transitional probabilities in nonhuman primates. *Cognitive Science*, *46*(4). https://doi.org/10.1111/cogs.13121

Rey, A., Jacobs, A. M., Schmidt-Weigand, F., & Ziegler, J. C. (1998). A phoneme effect in visual word recognition. *Cognition*, *68*(3), B71–B80. https://doi.org/10.1016/S0010-0277(98)00051-1

Rey, A., Minier, L., Malassis, R., Bogaerts, L., & Fagot, J. (2019). Regularity extraction across species: Associative learning mechanisms shared by human and non-human primates. *Topics in Cognitive Science*, *11*(3), 573–586. https://doi.org/10.1111/tops.12343

Rey, A., & Schiller, N. O. (2005). Graphemic complexity and multiple print-to-sound associations in visual word recognition. *Memory & Cognition*, *33*(1), 76–85. https://doi.org/10.3758/BF03195298

Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition*, *75*, B1–B12.

Robinet, V., Lemaire, B., & Gordon, M. B. (2011). MDLChunker: A MDL-Based cognitive model of inductive learning. *Cognitive Science*, *35*, 1352–1389. https://doi.org/10.1111/j.1551-6709.2011.01188.x

Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science*, *37*(7), 1290–1320. https://doi.org/10.1111/cogs.12050

Ryland Williams, J., Lessard, P. R., Desu, S., Clark, E. M., Bagrow, J. P., Danforth, C. M., & Sheridan Dodds, P. (2015). Zipfs law holds for phrases, not words. *Scientific Reports*, *5*, 1–7. https://doi.org/10.1038/srep12209

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Scarf, D., Smith, C., Jaswal, V., Magnuson, J., & Terrace, H. (2018). Chunky Monkey? The spontaneous temporal chunking of simultaneous chains by Humans (Homo Sapiens) and rhesus monkeys (Macaca mulatta). In *Studies of Rhesus Monkeys and their Behaviors* (pp. 39–58).

Slone, L. K., & Johnson, S. P. (2018). When learning goes beyond statistics: Infants represent visual sequences in terms of chunks. *Cognition*, *178*, 92–102. https://doi.org/10.1016/j.cognition.2018.05.016

Smalle, E. H. M., Bogaerts, L., Simonis, M., Duyck, W., Page, M. P. A., Edwards, M. G., & Szmalec, A. (2016). Can chunk size differences explain developmental changes in lexical learning? *Frontiers in Psychology*, *6*, 1–14. https://doi.org/10.3389/fpsyg.2015.01925

Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Animal Cognition*, *18*(3), 733–745. https://doi.org/10.1007/s10071-015-0840-x

Stahl, A. E., & Feigenson, L. (2014). Social knowledge facilitates chunking in infancy. *Child Development*, *85*(4), 1477–1490. https://doi.org/10.1111/cdev.12217

St-Louis, M.-È., Hughes, R. W., Saint-Aubin, J., & Tremblay, S. (2019). The resilience of verbal sequence learning: Evidence from the Hebb repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 17–25. https://doi.org/10.1037/xlm0000566

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86–132. https://doi.org/10.1016/j.cogpsych.2004.06.001

Tosatto, L., Bonafos, G., Melmi, J.-B., & Rey, A. (2022). Detecting non-adjacent dependencies is the exception rather than the rule. *PLOS ONE*, *17*(7), e0270580. https://doi.org/10.1371/journal.pone.0270580

Tosatto, L., Fagot, J., Nemeth, D., & Rey, A. (2022). The evolution of chunks in sequence learning. *Cognitive Science*, *46*(4). https://doi.org/10.1111/cogs.13124

Tosatto, L., Fagot, J., Nemeth, D., & Rey, A. (in press). Chunking as a function of sequence length. *Animal Cognition*.

Tosatto, L., Fagot, J., & Rey, A. (2023a). Sequence learning and chunk stability in Guinea baboons (Papio papio). *Revue de Primatologie*, *14*, 69–81.

Tosatto, L., Fagot, J., & Rey, A. (2023b). The dynamics of chunking in humans (Homo sapiens) and Guinea baboons (Papio papio). *Journal of Comparative Psychology*. https://doi.org/10.1037/com0000336

Tovar, Á. E., & Westermann, G. (2017). A neurocomputational approach to trained and transitive relations in equivalence classes. *Frontiers in Psychology*, *8*, 1848. https://doi.org/10.3389/fpsyg.2017.01848

Tovar, Á. E., & Westermann, G. (2023). No need to forget, just keep the balance: Hebbian neural networks for statistical learning. *Cognition*, *230*, 105176. https://doi.org/10.1016/j.cognition.2022.105176

Tovar, Á. E., Westermann, G., & Torres, A. (2018). From altered synaptic plasticity to atypical learning: A computational model of Down syndrome. *Cognition*, *171*, 15–24. https://doi.org/10.1016/j.cognition.2017.10.021

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*, 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Van Der Kant, A., Männel, C., Paul, M., Friederici, A. D., Höhle, B., & Wartenburger, I. (2020). Linguistic and non-linguistic non-adjacent dependency learning in early development. *Developmental Cognitive Neuroscience*, *45*, 100819. https://doi.org/10.1016/j.dcn.2020.100819

Verwey, W. B. (2001). Concatenating familiar movement sequences: The versatile cognitive processor. *Acta Psychologica*, *106*(1–2), 69–95. https://doi.org/10.1016/S0001-6918(00)00027-5

Verwey, W. B., Abrahamse, E. L., Ruitenberg, M. F. L., Jiménez, L., & de Kleine, E. (2011). Motor skill learning in the middle-aged: Limited development of motor chunks and explicit sequence knowledge. *Psychological Research*, *75*(5), 406–422. https://doi.org/10.1007/s00426-011-0320-0

Vilkaitė, L. (2016). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1632–1642. https://doi.org/10.1037/xlm0000259

Watson, S. K., Burkart, J. M., Schapiro, S. J., Lambeth, S. P., Mueller, J. L., & Townsend, S. W. (2020). Nonadjacent dependency processing in monkeys, apes, and humans. *Science Advances*, *6*(43), eabb0725. https://doi.org/10.1126/sciadv.abb0725

Weyers, I., Männel, C., & Mueller, J. L. (2022). Constraints on infants' ability to extract non-adjacent dependencies from vowels and consonants. *Developmental Cognitive Neuroscience*, *57*, 101149. https://doi.org/10.1016/j.dcn.2022.101149

Wickelgren, W. A. (1967). Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology*, *19*(2), 97–102. https://doi.org/10.1080/14640746708400077

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., van der Kant, A., Smith, K., & Rey, A. (2020). Non-adjacent dependency learning in humans and

other animals. *Topics in Cognitive Science*, *12*(3), 843–858. https://doi.org/10.1111/tops.12381

Wymbs, N. F., Bassett, D. S., Mucha, P. J., Porter, M. A., & Grafton, S. T. (2012). Differential recruitment of the sensorimotor putamen and frontoparietal cortex during motor chunking in humans. *Neuron*, *74*(5), 936–946. https://doi.org/10.1016/j.neuron.2012.03.038

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Chapter 6: General discussion

"*Words used together fuse together*"

Bybee, Joan (2003)

As mentioned in the introduction to this thesis, language acquisition has been a subject of great interest in the field of psychology over the years. Contrary to the classical and long-standing idea that the building blocks of language are words and rules, usage-based approaches suggest that constructions are the backbone of language, in particular multiword sequences. Since the 2000s, corpus analyses have revealed that multiword sequences are ubiquitous in language and constitute a large part of natural speech. Furthermore, a large body of empirical research has shown that multiword sequences play a key role in language processing. While most of these studies have focused on their processing, only a limited number have investigated how these sequences are acquired in real time. Therefore, the overall objective of this thesis was to investigate the extraction and learning dynamics of multiword sequences, as well as to examine the underlying learning mechanisms.

To bridge this gap we conducted three independent series of experimental studies. We first investigated the learning dynamics of different types of multiword sequences and how repetition affects each word within the repeated sequence (Chapter 2). Secondly, we investigated the influence of repetition spacing on their acquisition (Chapter 3). In these two chapters, we also examined the associative mechanism underlying sequence learning, namely chunking. We then aimed to disentangle memory-specific from language-related components of chunking by investigating the influence of repetition-spacing on sequence

learning in baboons (Chapter 4). Finally, we discussed how chunking (Chapter 5) can account for sequence learning and for the experimental effects that I have uncovered in this thesis.

6. Summary of findings

Study 1 (Chapter 2) employed the newly developed Hebb lexical decision task, in which a word (Experiments 1, 3 and 4) or pseudoword (Experiment 2) triplet was repeated in a noisy stream of random words and pseudowords. We found that participants were able to learn the novel multiword sequence after about four to five repetitions. We also found that response times (RTs) for the three words in the sequence followed a distinct trajectory. Indeed, RTs for the unpredictable first position in the triplet decreased over repetitions, which we termed the **repetition effect**, but more slowly than for words appearing in the predictable second and third positions in the repeated triplet (i.e., the **predictability effect**). Interestingly, learning, both implicit and explicit, also varied as a function of the type of sequence (i.e., unrelated words, pseudowords, semantically related words, or idioms), with semantically related words and idioms being processed faster and recalled better.

In **Study 2** (Chapter 3), we sought to investigate the influence of repetition spacing on the acquisition of multiword sequences and on the development of their memory trace. To this end, we used the Hebb lexical decision task, but, unlike Study 1, in which each repetition of the triplet was separated by three to six fillers, here, the triplet was systematically repeated every 4th, 7th, 10th, 20th, 30th or 60th trial, depending on the spacing condition. Our results showed that learning was still possible when the target triplets were separated by as many as 60 fillers, and that the forgetting curve appeared to follow a power law. In addition, we replicated the repetition and predictability effects

observed in Study 1, both of which were affected by repetition spacing. In fact, the repetition effect increased while the predictability effect decreased as spacing increased.

In **Study 3** (Chapter 4), we tested a group of Guinea baboons in a visuo-motor pointing task which combined the Serial Reaction Time task and the Hebb repetition paradigm. The aim of this study was to test the development of a sequential memory trace in the absence of the use of verbal recoding strategies. As in the human task, we systematically manipulated the interval between two repetitions of a target sequence by varying the number of random sequences in between. Thus, a sequence consisting of three points was systematically repeated every 1, 3, 6 or every trial, depending on the spacing condition. We found that baboons can learn repeated visuomotor sequences, even when the repetitions are separated by six random sequences. Similar to human results, these findings showed that learning decreased with increasing spacing and that forgetting followed a non-linear trajectory, which was best fitted by a logarithmic function. Taken together, Studies 2 and 3 suggest that memory traces of sequential information are long-lived.

Finally, in **Chapter 5**, we summarised the results of the present thesis and of recent empirical work on the main phenomena that have been consistently reported in the literature on chunking mechanisms in both humans and non-human primates: the predictability effect, the influence of repetition spacing, and chunk size limits. Importantly, we showed that current chunk-based models cannot account for these effects and argued that recent neurocomputational models based on associative and Hebbian learning may be more appropriate for describing and understanding the nature of chunking mechanisms.

6.1. The effect of repetition in language acquisition

Over the last few decades, psycholinguistic research has demonstrated that learners are highly sensitive to event frequency at different levels of language such as phonology, language comprehension and production, reading, and syntax. In fact, people learn and process high-frequency sequences and regular linguistic patterns more easily. Accordingly, usage-based approaches emphasise the importance of language use and repetition in language acquisition. As Bybee (2010) notes, repetition is what triggers chunking, which allows learners who encounter a new word or sequence of words to develop a memory trace and, after a sufficient number of repetitions, to consolidate it in memory. The more times we process an item, the stronger our memory for it and the easier it is to access. Frequency is therefore thought to play a key role in language acquisition (Ellis, 2002).

As research with infants shows, word learning occurs very rapidly after only a few repetitions. For instance, 13- to 16-month-old infants are able to perform word mapping after only four repetitions (Kay-Raining Bird & Chapman, 1998), and 9- to 11-year-old children are also able to learn new words after only four repetitions (Tamura et al., 2017). Similarly, it has been suggested that the acquisition of multiword sequences follows a similar trajectory, with four to five repetitions of a multiword sequence being sufficient to observe a change in participants' reading behaviour (Conklin & Carrol, 2020; Pellicer-Sánchez et al., 2022). Consistent with these findings, our results from **Study 1** also suggest that approximately four repetitions are sufficient to acquire a novel three-word multiword sequence (hereafter triplet). Indeed, our broken-stick analyses revealed a rapid decrease in RTs for the second and third predictable words in the triplet up to the fourth repetition. Additionally, the emergence of a difference between the first unpredictable and predictable words also occurred after approximately four repetitions.

The learning trajectories of the multiword sequences in **Study 2** seem to follow a similar dynamic.

Interestingly, the learning dynamics for the triplets of pseudowords were very similar to those of words, although when we looked in detail at the evolution of RTs for each item within the triplet, we observed that repetition differentially affected the processing of items already encoded in memory (i.e., words) compared to novel items (i.e., pseudowords), with pseudowords benefiting much more from repetition. This is probably related to the fact that previously encoded words are already familiar to participants and therefore benefit less from recent exposure, unlike pseudowords that have never been seen before. Such a phenomenon is particularly evident when comparing high-frequency to low-frequency words in lexical decision tasks, where low-frequency words have been shown to benefit more from repetition than high-frequency words, for which performance is closer to a ceiling (Duchek & Neely, 1989; Ober & Shenaut, 2014; Scarborough et al., 1977).

While our findings show that repetition is essential for the development of memory traces, they also suggest that a high frequency of exposure does not necessarily imply a gain in processing speed. Indeed, our results from **Study 1** show that the processing of the target multiword sequence benefits greatly from repetition during the first few exposures, but the amount of processing gain decreases as a function of frequency, as illustrated by the different regimes of the broken-stick analyses, almost reaching a ceiling around the twentieth repetition. These results are consistent with the power law of learning (Anderson, 1982; Ellis & Schmidt, 1997), which states that the effects of learning are strongest in the early stages of learning, but decrease with practice until an asymptote is reached (Ellis, 2002; Logan, 1990). Note that chunking has also been shown to be subject to the power law of practice (Bybee, 2010; Newell & Rosenbloom, 1981).

6.2. Repetition and predictability effects

Our results from **Studies 1** and **2** also shed light on the different learning dynamics that each word in the repeated triplet follows throughout the task. On the one hand, we showed that participants tended to process the first unpredictable word in the triplet faster as the task progressed (i.e., **the repetition effect**). However, this mere repetition effect was rather limited, with a processing gain of about 80 ms after 45 repetitions. On the other hand, participants responded much faster to the second and third predictable words in the triplet than to the first unpredictable word (i.e., **the predictability effect**). In fact, the processing advantage for predictable words was about 300 ms. This processing advantage can be attributed to the fact that the second and third predictable words benefit from the preceding word, which systematically occurs before them and helps participants to anticipate and predict the next item. Thus, the second word benefits from the systematic occurrence of the first word, whereas the third word benefits from the occurrence of both the first and second words in the triplet.

These findings are in line with previous studies showing that frequently encountered words are processed more efficiently (Brysbaert et al., 2018; Stark & McClelland, 2000), as well as studies showing a processing advantage for predictable items embedded in a repeated sequence (Rey et al., 2020; H. S. Wang et al., 2023). Although similar results have been reported previously, the originality of this work lies in the fact that we were able to study the evolution of these two effects simultaneously, thus providing an overview of the learning dynamics involved in the acquisition of multiword sequences.

Furthermore, our results from **Study 3**, in which we test a group of Guinea baboons in a Hebb visuo-motor pointing task, together with the review presented in **Chapter 5**, suggest that these effects are not specific to language. Rather, they are a common feature

of the chunking mechanisms operating in sequence learning and rely on basic associative processes shared by both species (Conway, 2020; Wilson et al., 2017).

6.3. The influence of repetition spacing in sequence learning

One of the aims of this thesis was to investigate the influence of repetition spacing on the learning dynamics of multiword sequences. Our results from **Study 2** indicate that multiword sequence learning is still possible at large repetition spacing (i.e., 60), and that the overall learning rate of a repeated sequence slows with increasing spacing. However, the influence of repetition spacing differed between the repetition and predictability effects. On the one hand, words in the first unpredictable positions seemed to benefit from spacing, with the repetition effect being greater for wider spacings. Our interpretation of this finding was that participants increasingly expected and anticipated the occurrence of the repeated triplet, resulting in faster responses. On the other hand, the processing advantage associated with the predictability effect decreased with increasing spacing. A potential twofold explanation for this finding is that, with greater spacing, the associations between the words of the triplet take longer to consolidate in memory, and concurrently, the same memory trace gradually disappears from memory as the sequence is not encountered repeatedly enough (Perruchet & Vinter, 1998). Evidence for this is provided by the fact that participants recalled the repeated triplet less accurately as the spacing increased. Furthermore, the decrease in the predictability effect followed a power law, which is classically associated with a weakening of memory traces and hence lower memory performance (Kahana & Adler, 2017; Wixted, 2022).

Similarly, results from **Study 3** showed that baboons were able to learn sequences of three locations even at the largest spacing, and that the learning rate also decreased as a function of spacing. Like in humans, the predictability effect decreased with increasing

spacing, but its course was better fitted by a logarithmic function. As mentioned in **Chapter 4**, this is not surprising as both functions have been shown to be highly correlated and are usually observed during implicit learning (Radvansky et al., 2024). It is worth noting, however, that recent studies suggest that the memory performance is not systematically captured by these functions. Instead, the best-fitting function varies with the complexity and meaningfulness of the material, the degree of learning, and the phase of retention at which the trace is located (Fisher & Radvansky, 2019; Radvansky et al., 2022).

6.4. Does repetition always result in the same trajectory for the development of memory traces?

Thus far, I have mainly emphasised the role of repetition in the learning of multiword sequences. However, according to usage-based approaches, other distributional information contained in the input, as well as the meaning of the constructions, should also influence their acquisition and processing. In line with this assumption, our findings from **Study 1** revealed that idioms were processed faster than semantically related words, which in turn were processed faster than the unrelated words. As discussed in **Chapter 2**, this can be attributed to pre-existing associations between words that facilitated the processing of both semantically related words and idioms. Additionally, idioms may have benefitted from their meaningfulness. Indeed, prior research have demonstrated that multiword sequences with a higher degree of meaningfulness are easier to process and result in lower RTs and reading times compared to control sequences (Chantavarin et al., 2022; Jolsvai et al., 2020).

It is noteworthy that the participants' ability to recall the repeated triplet was also likely influenced by these factors, with idioms being recalled more accurately than

semantically related words, which were also recalled better than unrelated words. Interestingly, the recall performance seemed to be related to the processing speed of the sequence, but not to the probability of learning the sequence, as the predictability effect was substantial across conditions. This suggests that a memory trace for the triplet has been created, albeit only partially retrievable. The results of **Study 2** provide further evidence for this, as both recall performance and predictability decreased with spacing, yet learning was still observed across all spacing conditions. Furthermore, an examination of the recall of each triplet at the word level reveals that recall performance of the third word appears to be relatively preserved, whereas recall of the second and first words appears to decline with increasing spacing (see Table 1).

Table 1

Number of words correctly recalled for each position in the triplet and for each spacing condition.

	Position 1	Position 2	Position 3
4F	69	61	73
7F	52	52	68
10F	54	57	69
20F	46	52	68
30F	31	43	67
60F	35	42	62

Taken together, it seems difficult to explain these results solely on the basis of the chunk-based models I have examined in **Chapters 1** and **5**, which are the traditional models cited in the field of language acquisition. In these models, learning depends only on the frequency of co-occurrence of the elements constituting the sequence, without taking into account factors that have been shown to influence this process, such as the semantic relationship between elements, meaningfulness, idiomaticity or familiarity (e.g., Carrol & Conklin, 2020; Fioravanti & Siyanova-Chanturia, 2024; Jolsvai et al., 2020).

Consequently, two sequences with the same frequency are always encoded and processed in the same way by these models, despite the inherent differences in the properties of the words and the sequences themselves. Furthermore, it is unclear how these models could account for the reduced recall performance for only some of the words in the triplet. Statistical learning models based on the principles of Hebbian learning could perhaps provide clues to answer this problem. Since each association between words within the sequence is modelled independently, it is possible to assign different weights to each association to account for the likely uneven encoding of each word.

6.5. Limitations of the Hebb lexical decision task

The main limitation of this research concerns the Hebb lexical decision task that we developed. Indeed, this task, which was used throughout the thesis, allowed us to study the acquisition dynamics of multiword sequences in real time, with very fine measurements for each word in the sequence. However, this task is far from mimicking a real reading situation in which multiword sequences are relatively spaced from one another and occur much less frequently. Furthermore, this task seems unsuitable for studying the acquisition of multiword sequences using naturalistic stimuli, as it would imply using real sentences for the targets and for the fillers, which is impossible given that a lexical decision task requires the presence of pseudowords in order to be performed.

A second limitation concerns the size of the sequences we can use in the Hebb lexical decision task. Virtually, we could use multiword sequences of any length, but this might lead participants to learn that there are a certain number of words that are repeated without properly processing each word. Thus, the task would no longer be assessing the acquisition of the sequence, but rather the extraction of the size of the regularity presented in the task.

A third limitation relates to lexical decision itself as a proxy for the study of language acquisition. Indeed, this task requires the involvement of multiple components of the language processing system, such as orthographic, phonological and semantic components, in order to make a lexical decision (Christiansen, 2019). This task therefore involves many more processes than a normal reading situation or conversation where individuals do not constantly analyse each word in the sequence for comprehension and production (Ferreira et al., 2002; Goldberg & Ferreira, 2022).

One possible solution to these problems, it is to adapt the Hebb lexical decision task into a naming task to avoid decision making, but also to ensure that participants process each word. In order to perform the task correctly, participants would have to read each word. Despite all these shortcomings, I strongly believe that the Hebb lexical decision task remains an interesting paradigm because of its potential to be adapted into different types of tasks that may permit the study of sequence learning and chunking in different modalities.

6.6. Future perspectives

Our first perspective stems from the fact that we have studied the acquisition of multiword sequences in a well-controlled environment that does not correspond to the experience of real-life language use. Future research should use other tasks and methods that allow the use of more ecological stimuli, while maintaining the level of detail we achieved with the Hebb lexical decision task. One particularly interesting method that may be suitable for this purpose is eye-tracking, which allows the study of participants' eye movements during naturalistic reading, while providing a range of measures that tap into different stages of language processing, such as lexical access, initial integration of information, text comprehension and recovery from processing difficulties (Siyanova-

Chanturia, 2013). As mentioned in the introduction, this method has mostly been used to study the processing of already encoded multiword sequences. The idea would be to use protocols such as Conklin and Carrol's (2020) with different types of multiword sequences to get a more complete picture of how they are acquired.

A second perspective relates to the discussion of chunk size in Chapter 5. In that chapter, we pointed out that defining the elements that form a chunk can be difficult because it is unclear what should be considered as its basic unit. For instance, when analysing the phrase "We haven't got enough", we can decompose it into four words, into six syllables, or into 14 phonemes, making it difficult to know what is really represented in memory. Therefore, an essential issue for future research will be to develop a method for determining what the basic elements of a chunk are. This would allow us to answer questions such as how chunk hierarchies are built, how chunk size is determined by our cognitive system, and whether there is a limit to chunk size.

Finally, as discussed throughout this thesis and in more detail in Chapter 5, chunk-based models cannot fully account for all current findings, namely the **repetition** and **predictability** effects. An interesting avenue for future work would be to develop a computational model that can capture these phenomena. My collaborators and I are currently working on a new model that combines the well-known *Interactive Activation Model* (IA) developed by McClelland and Rumelhart (1981) with the Hebbian learning principles. On the one hand, the IA model simulates the general neural activation underlying letter and word recognition in humans by relying on a connectionist architecture consisting of three levels: visual features, letters and words. On the other hand, Hebbian learning posits that repeated processing of two items leads to the formation of a synaptic connection between the neuron populations encoding them, with this connection strengthening with further coactivation, and weakening otherwise. Since

we are interested in modelling the acquisition of multiword sequences, we only implemented the letter and word levels in our model. Therefore, the architecture of the model consists of a first **letter level** where each letter position in a word is represented by a "position" section containing 26 letter nodes. Depending on which word is presented to the model, a letter node is activated in each of the position sections. Letter activation is then propagated to the **word level**, which consists of all the words in the vocabulary. Here, each word node receives excitation or inhibition from the letter level, depending on whether its letter nodes are activated or not. Once a word node reaches a certain threshold, the word is recognised by the model. The time it takes for a word node to reach the threshold depends on the interaction between the resting state of the word (defined by its frequency) and its external activation. To simulate the **repetition effect**, each time a word is encountered, its resting state value increases, resulting in faster recognition the next time it is encountered. To simulate the **predictability effect**, when words are encountered together within certain time windows, they send excitation to each other, leading to increasingly faster recognition each time they are encountered together.

7. Conclusion

The three empirical studies and the review article presented in this thesis highlight the influence of repetition, of the nature of multiword sequences, and of repetition spacing on the acquisition of multiword sequences in real-time. Our results suggest that the learning dynamics observed in these studies are not language specific, but are a common feature of the chunking mechanisms that operate during sequence learning. However, current chunk-based models are unable to fully account for our findings. Taken together, this research demonstrates that sequential information is acquired rapidly, developing a transient memory trace, and that repetition plays a key role in its consolidation. It also

suggests that the lifetime of memory traces is potentially very long. In sum, these results raise important new questions for the field of language acquisition, and for the field of statistical learning and chunking more broadly.

Bibliographie

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *Linguistic Review*, *23*(3), 275–290. https://doi.org/10.1515/TLR.2006.011

Abu-Zhaya, R., Arnon, I., & Borovsky, A. (2022). Do children use multi-word information in real-time sentence comprehension? *Cognitive Science*, *46*(3). https://doi.org/10.1111/cogs.13111

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369–406.

Arcara, G., Lacaita, G., Mattaloni, E., Passarini, L., Mondini, S., Benincà, P., & Semenza, C. (2012). Is 'hit and run' a single word? The processing of irreversible binomials in neglect dyslexia. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00011

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth—Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107–129. https://doi.org/10.1080/15475441.2010.505489

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*(3), 349–371. https://doi.org/10.1177/0023830913484891

Arnon, I., & Cohen Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, *9*(3), 377–400. https://doi.org/10.1075/ml.9.3.01arn

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, *92*, 265–280. https://doi.org/10.1016/j.jml.2016.07.004

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82. https://doi.org/10.1016/j.jml.2009.09.005

Aslin, R. N. (2018). Statistical learning: A powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*, 1–12. https://doi.org/10.1002/wcs.1373.Statistical

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324. https://doi.org/10.1111/1467-9280.00063

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning of four-word combinations. *Psychological Science*, *19*(3), 241–248. https://doi.org/10.1111/j.1467-9280.2008.02075.x

Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics*, *47*(2). https://doi.org/10.1515/LING.2009.014

Brand, M., Giroux, I., Puijalon, C., & Rey, A. (2007). Syllable onsets are perceptual reading units. *Memory & Cognition*, *35*(5), 966–973. https://doi.org/10.3758/BF03193469

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*(1–2), 93–125. https://doi.org/10.1016/S0010-0277(96)00719-6

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711–733.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Cabiddu, F., Bott, L., Jones, G., & Gambi, C. (2023). CLASSIC utterance boundary: A chunking-based model of early naturalistic word segmentation. *Language Learning*, 1–34. https://doi.org/10.1111/lang.12559

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*(6), 843–873. https://doi.org/10.1207/s15516709cog2706_2

Cappelle, B., Shtyrov, Y., & Pulvermüller, F. (2010). Heating up or cooling up the brain? MEG evidence that phrasal verbs are lexical units. *Brain and Language*, *115*(3), 189–201. https://doi.org/10.1016/j.bandl.2010.09.004

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, *63*(1), 95–122. https://doi.org/10.1177/0023830918823230

Carrol, G., & Littlemore, J. (2020). Resolving figurative expressions during reading: The role of familiarity, transparency, and context. *Discourse Processes*, *57*(7), 609–626. https://doi.org/10.1080/0163853X.2020.1729041

Chantavarin, S., Morgan, E., & Ferreira, F. (2022). Robust processing advantage for binomial phrases with variant conjunctions. *Cognitive Science*, *46*(9). https://doi.org/10.1111/cogs.13187

Chater, N., McCauley, S. M., & Christiansen, M. H. (2016). Language as skill: Intertwining comprehension and production. *Journal of Memory and Language*, *89*, 244–254. https://doi.org/10.1016/j.jml.2015.11.004

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1959). A review of B. F. Skinner's Verbal Behavior. *Language*, *35*(1), 26–58. https://doi.org/10.4159/harvard.9780674594623.c6

Chomsky, N. (1980). Rules and representations. *The Behavioral and Brain Sciences*, *3*, 1–61.

Chomsky, N. (2014). *The minimalist program*. MIT Press.

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, 1–14. https://doi.org/10.1111/tops.12332

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*(2–3), 221–268. https://doi.org/10.1080/016909698386528

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science*, *9*(3), 542–551. https://doi.org/10.1111/tops.12274

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, 1–72. https://doi.org/10.1017/S0140525X1500031X

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*(10), 406–416. https://doi.org/10.1016/S1364-6613(98)01232-7

Conklin, K., & Carrol, G. (2020). Words go together like 'bread and butter': The rapid, automatic acquisition of lexical patterns. *Applied Linguistics*, *42*(3), 1–23. https://doi.org/10.1093/applin/amaa034

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*(1), 72–89. https://doi.org/10.1093/applin/amm022

Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience and Biobehavioral Reviews*, *112*, 279–299. https://doi.org/10.1016/j.neubiorev.2020.01.032

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Dąbrowska, E. (2015). What exactly is Universal Grammar, and has anyone seen it? *Frontiers in Psychology*, *6*(June), 1–17. https://doi.org/10.3389/fpsyg.2015.00852

Duchek, J. M., & Neely, J. H. (1989). A dissociative word-frequency X levels-of-processing interaction in episodic recognition and lexical decision tasks. *Memory & Cognition*, *17*(2), 148–162. https://doi.org/10.3758/BF03197065

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, N. C., & Ogden, D. C. (2017). Thinking about multiword constructions: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science*, *9*(3), 604–620. https://doi.org/10.1111/tops.12256

Ellis, N. C., & Schmidt, R. (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition*, *19*(2), 145–171. https://doi.org/10.1017/S0272263197002027

Elsherif, M. M., Catling, J. C., & Frisson, S. (2020). Two words as one: A multi-naming investigation of the age-of-acquisition effect in compound-word processing. *Memory & Cognition*, *48*(4), 511–525. https://doi.org/10.3758/s13421-019-00986-6

Emerson, S. N., & Conway, C. M. (2023). Chunking versus transitional probabilities: Differentiating between theories of statistical learning. *Cognitive Science*, *47*(5), e13284. https://doi.org/10.1111/cogs.13284

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66–108. https://doi.org/10.1016/j.dr.2015.05.002

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk*, *20*(1), 29–62. https://doi.org/10.1515/text.1.2000.20.1.29

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254–260. https://doi.org/10.1111/j.1467-9280.2007.01885.x

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Fioravanti, I., & Siyanova-Chanturia, A. (2024). Eye movements in the investigation of different properties of multi-word expressions: A systematic review. *Research*

Methods in Applied Linguistics, *3*(1), 100092. https://doi.org/10.1016/j.rmal.2023.100092

Fisher, J. S., & Radvansky, G. A. (2019). Linear forgetting. *Journal of Memory and Language*, *108*, 104035. https://doi.org/10.1016/j.jml.2019.104035

Fonollosa, J., Neftci, E., & Rabinovich, M. (2015). Learning of chunking sequences in cognition and behavior. *PLOS Computational Biology*, *11*(11), e1004592. https://doi.org/10.1371/journal.pcbi.1004592

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614–636. https://doi.org/10.1037/a0025255

Frost, R. L. A., Jessop, A., Durrant, S., Peter, M. S., Bidgood, A., Pine, J. M., Rowland, C. F., & Monaghan, P. (2020). Non-adjacent dependency learning in infancy, and its link to language development. *Cognitive Psychology*, *120*, 101291. https://doi.org/10.1016/j.cogpsych.2020.101291

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70–74. https://doi.org/10.1016/j.cognition.2015.11.010

Gobet, F. (2017). Entrenchment, Gestalt formation, and chunking. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 245–267). American Psychological Association. https://doi.org/10.1037/15969-012

Gobet, F., Lane, P., Croker, S., Cheng, P., Jones, G., Oliver, I., & Pine, J. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*(6), 236–243. https://doi.org/10.1016/S1364-6613(00)01662-4

Gobet, F., Lloyd-Kelly, M., & Lane, P. C. R. (2016). What's in a name? The multiple meanings of "Chunk" and "Chunking". *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00102

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.

Goldberg, A. E., & Ferreira, F. (2022). Good-enough language production. *Trends in Cognitive Sciences*, *26*(4), 300–311. https://doi.org/10.1016/j.tics.2022.01.005

Goldberg, A. E., & Herbst, T. (2021). The nice-of-you construction and its fragments. *Linguistics*, *59*(1), 285–318. https://doi.org/10.1515/ling-2020-0274

Goldberg, A. E., & Suttle, L. (2010). Construction grammar. *WIREs Cognitive Science*, *1*(4), 468–477. https://doi.org/10.1002/wcs.22

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5).

Hanna, J., Cappelle, B., & Pulvermüller, F. (2017). Spread the word: MMN brain response reveals whole-form access of discontinuous particle verbs. *Brain and Language*, *175*, 86–98. https://doi.org/10.1016/j.bandl.2017.10.002

Havron, N., & Arnon, I. (2020). Starting big: The effect of unit size on language learning in children and adults. *Journal of Child Language*, *48*(2), 244–260. https://doi.org/10.1017/S0305000920000264

Hernández, M., Costa, A., & Arnon, I. (2016). More than words: Multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience*, *31*(6), 785–800. https://doi.org/10.1080/23273798.2016.1152389

Ibbotson, P. (2013). The scope of usage-based theory. *Frontiers in Psychology*, *4*, 1–15. https://doi.org/10.3389/fpsyg.2013.00255

Isbilen, E. S., & Christiansen, M. H. (2018). Chunk-based memory constraints on the cultural evolution of language. *Topics in Cognitive Science*, *12*, 713–726. https://doi.org/10.1111/tops.12376

Isbilen, E. S., Frost, R. L. A., Monaghan, P., & Christiansen, M. H. (2022). Statistically based chunking of nonadjacent dependencies. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001207

Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2022). Individual differences in artificial and natural language statistical learning. *Cognition*, *225*, 105123. https://doi.org/10.1016/j.cognition.2022.105123

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science*, *44*(7). https://doi.org/10.1111/cogs.12848

Ivanchenko, D., Rifai, K., Hafed, Z. M., & Schaeffel, F. (2021). A low-cost, high-performance video-based binocular eye tracker for psychophysical research. *Journal of Eye Movement Research*, *14*(3). https://doi.org/10.16910/jemr.14.3.3

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0033202

Jeong, H., van den Hoven, E., Madec, S., & Bürki, A. (2021). Behavioral and brain responses highlight the role of usage in the preparation of multiword utterances for production. *Journal of Cognitive Neuroscience*, *33*(11), 2231–2264. https://doi.org/10.1162/jocn_a_01757

Jiang, S., Jiang, X., & Siyanova-Chanturia, A. (2020). The processing of multiword expressions in children and adults: An eye-tracking study of Chinese. *Applied Psycholinguistics*, *41*(4), 901–931. https://doi.org/10.1017/S0142716420000296

Jiang, S., & Siyanova–Chanturia, A. (2023). The processing of multiword expressions in L1 and L2 Chinese: Evidence from reaction times and eye movements. *The Modern Language Journal*, modl.12846. https://doi.org/10.1111/modl.12846

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345. https://doi.org/10.1111/j.1467-7687.2009.00886.x

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2020). Meaningfulness beats frequency in multiword chunk processing. *Cognitive Science*, *44*(10), e12885. https://doi.org/10.1111/cogs.12885

Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology*, *98*, 1–21. https://doi.org/10.1016/j.cogpsych.2017.07.002

Kahana, M. J., & Adler, M. (2017). *Note on the power law of forgetting*. Neuroscience. https://doi.org/10.1101/173765

Kay-Raining Bird, E., & Chapman, R. S. (1998). Partial representations and phonological selectivity in the comprehension of 13- to 16-month-olds. *First Language*, *18*(52), 105–127. https://doi.org/10.1177/014272379801805204

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171–184. https://doi.org/10.1037/a0025405

Kyriacou, M., Conklin, K., & Thompson, D. (2020). Passivizability of idioms: Has the wrong tree been barked up? *Language and Speech*, *63*(2), 404–435. https://doi.org/10.1177/0023830919847691

Kyriacou, M., Conklin, K., & Thompson, D. (2021). When the idiom advantage comes up short: Eye-tracking canonical and modified idioms. *Frontiers in Psychology*, *12*, 675046. https://doi.org/10.3389/fpsyg.2021.675046

Kyriacou, M., Conklin, K., & Thompson, D. (2022). Ambiguity resolution in passivized idioms: Is there a shift in the most likely interpretation? *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*. https://doi.org/10.1037/cep0000300

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Li, H., Warrington, K. L., Pagán, A., Paterson, K. B., & Wang, X. (2021). Independent effects of collocation strength and contextual predictability on eye movements in reading. *Language, Cognition and Neuroscience*, *36*(8), 1001–1009. https://doi.org/10.1080/23273798.2021.1922726

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, *30*(2), 333–370. https://doi.org/10.1017/S0305000903005592

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, *22*(1), 1–35. https://doi.org/10.1016/0010-0285(90)90002-L

Malassis, R., Rey, A., & Fagot, J. (2018). Non-adjacent dependencies processing in human and non-human primates. *Cognitive Science*, *42*(5), 1677–1699. https://doi.org/10.1111/cogs.12617

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, *122*(3), 346–362. https://doi.org/10.1016/j.cognition.2011.11.003

Matthews, D., & Bannard, C. (2010). Children's production of unfamiliar word sequences is predicted by positional variability and latent classes in a large sample of child-directed speech. *Cognitive Science*, *34*(3), 465–488. https://doi.org/10.1111/j.1551-6709.2009.01091.x

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134. https://doi.org/10.1111/j.1467-7687.2007.00653.x

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111. https://doi.org/10.1016/S0010-0277(01)00157-3

McCauley, S. M., & Christiansen, M. H. (2019a). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, *126*(1), 1–51. https://doi.org/10.1037/rev0000126

McCauley, S. M., & Christiansen, M. H. (2019b). Modeling children's early linguistic productivity through the automatic discovery and use of lexically-based frames. *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, 782–788.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, *20*(1), 67–85. https://doi.org/10.1016/S0022-5371(81)90312-1

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–56). Routledge.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2

Ober, B. A., & Shenaut, G. K. (2014). Repetition priming of words and nonwords in Alzheimer's disease and normal aging. *Neuropsychology*, *28*(6), 973–983. https://doi.org/10.1037/neu0000112

O'Grady, W., Dobrovolsky, M., & Aronoff, M. (Eds.). (1997). *Contemporary linguistics: An introduction* (3rd ed.). St. Martin's Press.

Pellicer-Sánchez, A., Siyanova-Chanturia, A., & Parente, F. (2022). The effect of frequency of exposure on the processing and learning of collocations: A comparison of first and second language readers' eye movements. *Applied Psycholinguistics*, *43*(3), 1–30. https://doi.org/10.1017/S014271642200011X

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244–247. https://doi.org/10.1016/j.cognition.2009.07.011

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009b). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685. https://doi.org/10.1111/j.1467-8624.2009.01290.x

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Science*, *11*(3), 520–535. https://doi.org/10.1111/tops.12403

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246–263. https://doi.org/10.1006/jmla.1998.2576

Perry, C. (2023). Graphemes are used when reading: Evidence from Monte Carlo simulation using word norms from mega-studies. *Quarterly Journal of Experimental Psychology*, *76*(2), 419–428. https://doi.org/10.1177/17470218221086533

Pinker, S. (1994). *The language instinct: How the mind creates language*. HarperCollins.

Pinker, S. (1998). Words and rules. *Lingua*, *106*, 219–242. https://doi.org/10.1016/S0024-3841(98)00035-7

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456–463. https://doi.org/10.1016/S1364-6613(02)01990-3

Pulido, M. F. (2021). Individual chunking ability predicts efficient or shallow L2 processing: Eye-tracking evidence from multiword units in relative clauses. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.607621

Radvansky, G. A., Doolen, A. C., Pettijohn, K. A., & Ritchey, M. (2022). A new look at memory retention and forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(11), 1698–1723. https://doi.org/10.1037/xlm0001110

Radvansky, G. A., Parra, D., & Doolen, A. C. (2024). Memory from nonsense syllables to novels: A survey of retention. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-024-02514-3

Redington, M., & Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends in Cognitive Sciences*, *1*(7), 273–281. https://doi.org/10.1016/S1364-6613(97)01081-4

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*(4), 425–469. https://doi.org/10.1207/s15516709cog2204_2

Rey, A., Fagot, J., Mathy, F., Lazartigues, L., Tosatto, L., Bonafos, G., Freyermuth, J., & Lavigne, F. (2022). Learning higher-order transitional probabilities in nonhuman primates. *Cognitive Science*, *46*(4). https://doi.org/10.1111/cogs.13121

Rey, A., Jacobs, A. M., Schmidt-Weigand, F., & Ziegler, J. C. (1998). A phoneme effect in visual word recognition. *Cognition*, *68*(3), B71–B80. https://doi.org/10.1016/S0010-0277(98)00051-1

Rey, A., Minier, L., Malassis, R., Bogaerts, L., & Fagot, J. (2019). Regularity extraction across species: Associative learning mechanisms shared by human and non-human primates. *Topics in Cognitive Science*, *11*(3), 573–586. https://doi.org/10.1111/tops.12343

Rey, A., & Schiller, N. O. (2005). Graphemic complexity and multiple print-to-sound associations in visual word recognition. *Memory & Cognition*, *33*(1), 76–85. https://doi.org/10.3758/BF03195298

Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition*, *75*, B1–B12.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, *81*(2), 149–169. https://doi.org/10.1016/S0010-0277(01)00132-9

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*(1), 172–196. https://doi.org/10.1006/jmla.2001.2839

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, *4*(2), 273–284. https://doi.org/10.1207/S15327078IN0402_07

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 1–17. https://doi.org/10.1037/0096-1523.3.1.1

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "Learning" in statistical learning: What does an online measure reveal about the assimilation of visual regularities? *Cognitive Science*, *42*, 692–727. https://doi.org/10.1111/cogs.12556

Siyanova-Chanturia, A. (2013). Eye-tracking and ERPs in multi-word expression research: A state-of-the-art review of the method and findings. *The Mental Lexicon*, *8*(2), 245–268. https://doi.org/10.1075/ml.8.2.06siy

Siyanova-Chanturia, A., Conklin, K., Ca, S., Kaan, E., & Heuven, W. J. B. V. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language*, *175*, 111–122. https://doi.org/10.1016/j.bandl.2017.10.004

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*(2), 251–272. https://doi.org/10.1177/0267658310382068

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase ' time and again' matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(3), 776–784. https://doi.org/10.1037/a0022531

Siyanova-Chanturia, A., & Janssen, N. (2018). Production of familiar phrases: Frequency effects in native speakers and second language learners. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 2009–2018. https://doi.org/10.1037/xlm0000562

Skarabela, B., Ota, M., O'Connor, R., & Arnon, I. (2021). 'Clap your hands' or 'take your hands'? One-year-olds distinguish between frequent and infrequent multiword phrases. *Cognition*, *211*, 104612. https://doi.org/10.1016/j.cognition.2021.104612

Smalle, E. H. M., & Möttönen, R. (2023). Cognitive development as a piece of the language learning puzzle. *Cognitive Science*, *47*(5), e13296. https://doi.org/10.1111/cogs.13296

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition*, *18*(3), 419–437. https://doi.org/10.1017/S1366728914000674

Stark, C. E. L., & McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(4), 945–972. https://doi.org/10.1037/0278-7393.26.4.945

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395–411. https://doi.org/10.1016/j.jecp.2014.06.003

Tamura, N., Castles, A., & Nation, K. (2017). Orthographic learning, fast and slow: Lexical competition effects reveal the time course of word learning in developing readers. *Cognition*, *163*, 93–102. https://doi.org/10.1016/j.cognition.2017.03.002

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160056. https://doi.org/10.1098/rstb.2016.0056

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. https://doi.org/10.1037/0012-1649.39.4.706

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, *3*(1), 73–100.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, *3*(1), 1–42. https://doi.org/10.1207/s15473341lld0301_1

Tomasello, M. (2001). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*(1–2), 61–82. https://doi.org/10.1515/cogl.2001.012

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tremblay, A., & Baayen, H. H. (2010). Holistic processing of regularfour-word sequences: A behavioral and ERP study of the effects ofstructure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). Continuum International Publishing Group.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*(2), 569–613. https://doi.org/10.1111/j.1467-9922.2010.00622.x

Tremblay, A., & Tucker, B. V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon*, *6*(2), 302–324. https://doi.org/10.1075/ml.6.2.04tre

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, *2*(10), 717–726. https://doi.org/10.1038/35094573

van Witteloostuijn, M., Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2019). Assessing visual statistical learning in early-school-aged children: The usefulness

of an online reaction time measure. *Frontiers in Psychology*, *10*(September), 1–16. https://doi.org/10.3389/fpsyg.2019.02051

Vilkaitė, L. (2016). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1632–1642. https://doi.org/10.1037/xlm0000259

Vilkaitė-Lozdienė, L. (2022). Do different morphological forms of collocations show comparable processing facilitation? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(9), 1328–1347. https://doi.org/10.1037/xlm0001130

Vilkaitė-Lozdienė, L., & Conklin, K. (2021). Word order effect in collocation processing. *The Mental Lexicon*, *16*(2–3), 362–396. https://doi.org/10.1075/ml.20022.vil

Vlach, H. A., & DeBrock, C. A. (2019). Statistics learned are statistics forgotten: Children's retention and retrieval of cross-situational word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 700–711. https://doi.org/10.1037/xlm0000611

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, *45*(6), 1611–1617. https://doi.org/10.1037/a0016134

Wang, F. H., Zevin, J., & Mintz, T. H. (2019). Successfully learning non-adjacent dependencies in a continuous artificial language stream. *Cognitive Psychology*, *113*(June), 101223. https://doi.org/10.1016/j.cogpsych.2019.101223

Wang, Y. (2019). A functional analysis of text-oriented formulaic expressions in written academic discourse: Multiword sequences vs. single words. *English for Specific Purposes*, *54*, 50–61. https://doi.org/10.1016/j.esp.2018.12.002

Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2010). Colliding cues in word segmentation: The role of cue strength and general cognitive processes. *Language and Cognitive Processes*, *25*(3), 402–422. https://doi.org/10.1080/01690960903212254

Wilson, B., Marslen-Wilson, W. D., & Petkov, C. I. (2017). Conserved sequence processing in primate frontal cortex. *Trends in Neurosciences*, *40*(2), 72–82. https://doi.org/10.1016/j.tins.2016.11.004

Wixted, J. T. (2022). The enigma of forgetting. *Proceedings of the National Academy of Sciences*, *119*(12), e2201332119. https://doi.org/10.1073/pnas.2201332119

Wray, A. (2002). Formulaic language and the lexicon. In *Language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, *32*, 231–254. https://doi.org/10.1017/S026719051200013X

Xu, Q., Chodorow, M., & Valian, V. (2023). How infants' utterances grow: A probabilistic account of early language development. *Cognition*, *230*, 105275. https://doi.org/10.1016/j.cognition.2022.105275

Yaneva, V., Taslimipoor, S., Rohanian, O., & Ha, L. A. (2017). Cognitive processing of multiword expressions in native and non-native speakers of english: Evidence from gaze data. *Computational and Corpus-Based Phraseology*, *10596*, 363–379. https://doi.org/10.1007/978-3-319-69805-2_26

Yi, W. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, *40*(4), 831–856. https://doi.org/10.1017/S0272263118000141

Yi, W., Lu, S., & Ma, G. (2017). Frequency, contingency and online processing of multiword sequences: An eye-tracking study. *Second Language Research*, *33*(4), 519–549. https://doi.org/10.1177/0267658317708009

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Table des figures

Figure 1. Frequency of n-grams plotted against their rank order, on logarithmic scales. Comparison of four-word sequences and single words. Figure adapted from Bannard and Matthews (2008)..... 27
Figure 2. Illustration of the stimuli used by Saffran et al. (1996) to study statistical learning in 8-month-old infants. (a) The list of syllables and three-syllable words. (b) The statistical structure of the words and part-words. Figure adapted from Aslin (2018)..... 32
Figure 3. Operations performed by PARSER at each processing step. Figure adapted from Perruchet and Vinter (1998). 36
Figure 4. Online processing of the utterance "the dog chased the cat". At each time step, the model calculates the backward transitional probabilities between adjacent words. If they exceed the average TP threshold (indicated by the backward arrow's position above the words), both words or chunks are grouped together. If they are below the running average, they are not grouped and are added to the chunkatory as separate chunks. Figure adapted from McCauley and Christiansen (2019a). 38
Figure 5. Experimental procedure of the Hebb lexical decision task. Panel A: items are presented one at a time in the centre of the screen. Participants have to classify each string as a word or a pseudoword. A repeated sequence of three words (e.g., W_1: "évier" – sink ; W_2: "raisin" – grape ; W_3: "flacon" – flask), always appearing in the same order, is interspersed with random filler words (W_R) or random filler pseudowords (PW_R). Words in blue belong to the repeated triplet. Panel B: one triplet of words ($W_1W_2W_3$) is repeated several times and a variable number of random words or pseudowords (W_R or PW_R) are presented between two repetitions of the triplet..... 44