

Recherche



Citez cet article : Heesen R, Hobaiter C, Ferrer-i-Cancho R, Semple S. 2019 Linguistic laws in chimpanzee gestural communication. *Proc. R. Soc. B* 286 : 20182900.
<http://dx.doi.org/10.1098/rspb.2018.2900>

Reçu : 20 décembre 2018
 Accepté : 16 janvier 2019

Catégorie de sujet :
 Comportement

Domaines d'activité :
 comportement, évolution

Mots-clés :
 lois linguistiques, compression, théorie de l'information, gestes, jeu

Auteur de la correspondance :
 Raphaela Heesen
 e-mail : raphaela.heesen@unine.ch

Le matériel complémentaire électronique est disponible en ligne à l'adresse suivante :
<https://dx.doi.org/10.6084/m9.figshare.c.4378208>.

Lois linguistiques dans la communication gestuelle des chimpanzés

Raphaela Heesen^{1,2}, Catherine Hobaiter³, Ramon Ferrer-i-Cancho⁴ and Stuart Semple¹

¹Centre de recherche en anthropologie évolutive, sociale et interdisciplinaire, Université de Roehampton, Londres SW15 4JD, Royaume-Uni

²Institut de Psychologie du Travail et des Organisations, Université de Neuchâtel, Rue Emile-Argand 11, 2000 Neuchâtel, Suisse

³School of Psychology and Neuroscience, University of St Andrews, St Andrews, Fife KY16 9JP, UK ⁴Complexity and Quantitative Linguistics Laboratory, Laboratory for Relational Algorithmics, Complexity, and Learning Research Group, Departament de Ciències de la Computació, Universitat Politècnica de Catalunya, 08034 Barcelona, Catalonia, Spain

© 2019 The Author(s) Published by the Royal Society. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Les études testant les lois linguistiques en dehors de la langue ont fourni des informations importantes sur l'organisation des systèmes biologiques. Par exemple, des modèles conformes à la loi de Zipf sur les abréviations (qui prédit une relation négative entre la longueur des mots et leur fréquence d'utilisation) ont été trouvés dans le comportement vocal et non vocal d'une série d'animaux, et des modèles conformes à la loi de Menzerath (selon laquelle les séquences plus longues sont composées d'éléments plus courts) ont été trouvés dans les séquences vocales des primates, ainsi que dans les gènes, les protéines et les génomes. Ces deux lois ont été associées à la compression, principe de la théorie de l'information qui consiste à minimiser la longueur du code. Nous présentons ici le premier test de ces lois dans la communication gestuelle animale. Au départ, nous n'avons pas trouvé la relation négative entre la durée du geste et la fréquence d'utilisation prédite par la loi de Zipf sur l'abréviation, mais cette relation a été observée dans des sous-ensembles spécifiques du répertoire. En outre, un modèle opposé à celui prédit a été observé dans un sous-ensemble de gestes - les signaux du corps entier. Nous avons constaté une corrélation négative entre le nombre et la durée moyenne des gestes dans les séquences, conformément à la loi de Menzerath. Ces résultats fournissent la première preuve que la compression sous-tend la communication gestuelle animale et mettent en évidence un point commun important entre la gestuelle des primates et le langage.

1. Introduction

L'étude des lois linguistiques - les modèles statistiques communs du langage humain - est une pierre angulaire de la linguistique quantitative [1,2]. Ces dernières années, des études ont commencé à explorer l'universalité des lois linguistiques au-delà de notre propre espèce, et ce travail a fourni des indications importantes sur les règles fondamentales d'organisation qui sous-tendent les systèmes d'information naturels. Plus particulièrement, l'exploration de deux de ces lois - la loi d'abréviation de Zipf et la loi de Menzerath - a prouvé que la compression, le principe de la théorie de l'information qui consiste à minimiser la longueur d'un code, est un principe universel non seulement du langage humain, mais aussi du comportement animal et d'une série d'autres systèmes d'information biologiques [3 - 5].

La loi de Zipf sur les abréviations prévoit une relation négative entre la longueur des mots et leur fréquence d'utilisation [6,7]. Elle est répandue dans un très grand nombre de langues humaines [8], dans les textes écrits (c'est-à-dire dans les systèmes d'écriture basés sur les caractères [9,10] et sur les lettres [8]), dans la parole [11] et dans la langue des signes [12]. Des modèles conformes à cette loi (c'est-à-dire une relation inverse entre l'ampleur du signal et la fréquence d'utilisation) ont également été documentés dans le comportement d'un certain nombre d'espèces animales : le répertoire vocal des macaques de Formose [5], les appels à courte distance des ouistitis communs [13], les appels sociaux des espèces de chauves-souris [14] et le comportement de surface non vocal des dauphins [15].

La loi de Menzerath prédit que "plus le tout est grand, plus ses constituants sont petits" et, dans le langage, elle s'applique à différentes échelles d'analyse : dans les mots comportant plus de syllabes, la longueur moyenne des syllabes est plus courte [16], et dans les phrases comportant plus de clauses, la longueur moyenne des clauses est plus courte [17]. Une relation négative entre la construction et la taille des constituants a été constatée dans les séquences vocales des géladas mâles [4] et des chimpanzés [18], et au niveau moléculaire - entre le nombre et la taille des chromosomes d'une espèce à l'autre [19], entre le nombre et la taille des exons dans les gènes [20], et entre le nombre et la taille des domaines dans les protéines [21].

Les explorations mathématiques indiquent que ces deux lois linguistiques reflètent la compression, et il a été proposé qu'il s'agisse d'un principe universel déterminant l'efficacité du codage [3,4,22]. On peut corroborer davantage l'effet de la compression dans le contexte de la loi de Zipf sur l'abréviation en vérifiant si la longueur moyenne du code est significativement faible dans les systèmes de signalisation qui suivent cette loi [3]. On a constaté que c'était le cas dans le langage humain et dans les systèmes animaux où cette loi s'applique [3]. En ce qui concerne la loi de Menzerath, une corroboration équivalente n'a pas encore été effectuée (chez l'homme ou chez d'autres espèces).

Bien que des preuves de la compression aient été trouvées dans une série de systèmes naturels, il est important d'élargir la gamme des modes de communication dans lesquels ce principe est étudié si l'on veut évaluer sa véritable portée. La communication gestuelle est un mode de signalisation important chez les primates anthropoïdes, y compris les humains [23], et il a été proposé qu'au cours de l'évolution humaine, la communication gestuelle a joué un rôle clé en facilitant l'émergence du langage parlé [23]. Les gestes sont définis comme des formes de communication non verbale impliquant des actions visibles, manuelles et corporelles ; ils se produisent généralement dans le cadre d'une communication à courte portée et sont utilisés dans un large éventail d'interactions sociales, notamment le jeu, le sexe, l'agression, l'allaitement et la toilette [24]. Parmi les systèmes gestuels des primates les mieux étudiés figure celui du chimpanzé, une espèce connue pour son vaste répertoire gestuel, avec des gestes donnés isolément ou combinés de manière flexible dans des séquences [25]. Les chimpanzés produisent 50 à 70 % des gestes de leur répertoire pendant le jeu social [24,25], ce qui fournit un contexte puissant pour tester la compression, tant au niveau des gestes individuels que des séquences de ces signaux. Nous analysons ici un ensemble complet de données sur les gestes de jeu recueillis auprès d'une communauté de chimpanzés sauvages, afin de tester la loi de Zipf sur l'abréviation des types de gestes individuels et la loi de Menzerath sur les séquences de gestes. Pour compléter ces deux tests de compression, nous vérifions également si la longueur moyenne du code est significativement faible dans les gestes individuels et les séquences, respectivement. Cette étude teste ces lois linguistiques dans un mode de signalisation animale dans lequel elles n'ont pas encore été étudiées, et fournit le premier test de la loi de Menzerath dans la signalisation gestuelle de n'importe quelle espèce, y compris l'homme. De plus, comme ces deux lois n'ont jamais été testées simultanément dans le même système en dehors de notre propre espèce, nos résultats fournissent de nouvelles informations sur les différents niveaux d'organisation des signaux auxquels la compression est appliquée.

peut être répandu dans des systèmes qui dépassent le langage humain.

2. Méthodes

(a) Site de l'étude et sujets

Nous avons effectué des observations sur les chimpanzés de la communauté de Sonso dans la réserve forestière de Budongo, en Ouganda. Au moment de la

Lors de l'étude, la communauté était composée de 81 membres identifiables. Nous avons défini les classes d'âge comme suit : nourrissons (0 - 4 ans), jeunes (5 - 9 ans), subadultes (femmes : 10 - 14 ans, hommes : 10 - 15 ans) et adultes (hommes : 10 - 15 ans). les adultes (femmes : plus ou moins 15 ans ; hommes : plus ou moins 16 ans).

(b) Collecte de données

Nous avons collecté des données au cours de quatre périodes de terrain - octobre 2007 - mars 2008 ; juin 2008 - janvier 2009 ; mai 2009 - août 2009 ; janvier 2011 - août 2011 - en utilisant un échantillonnage comportemental focal [26], avec des observations menées de 7h30 à 16h30. Nous avons enregistré des exemples de communication gestuelle pendant le jeu social à l'aide d'un Sony Handycam (DCR-HC-55). Le jeu social a été défini comme des situations où deux individus ou plus se sont engagés dans des activités de jeu indiquées par des signes de rire, des visages de jeu et des actions corporelles typiques telles que la lutte, la poursuite, les morsures de jeu ou les chatouilles [27].

(c) Codage

Au total, nous avons analysé 359 clips vidéo à la recherche de gestes de jeu répondant à au moins un des critères clés de la communication intentionnelle :

(i) sensibilité à l'état attentionnel du récepteur, (ii) attente de réponse, ou (iii) persistance du but [28]. Pour chacun de ces gestes, nous avons enregistré le type de geste (58 types ont été observés au total ; matériel électronique supplémentaire, S1), l'identité du signaleur, la durée du geste et le temps entre les gestes si les gestes ont été donnés en séquence.

(i) Mesurer la durée d'un geste

Nous avons mesuré la durée du geste en images, chaque image étant d'une durée de 0,04 s, en utilisant MPEG STREAMCLIP (v. Squared 5, 2012). Nous avons déterminé le début du geste comme le commencement du mouvement des parties du corps participant au processus gestuel. Nous avons enregistré la fin du geste soit comme l'arrêt des mouvements du corps créant le geste, soit comme le changement de position du corps si le geste reposait sur certains alignements du corps. Si le signaleur est resté dans la position du geste tout en commençant à jouer, nous l'avons considéré comme le point final du geste, car le geste ne répondait plus aux critères de la communication intentionnelle [28].

(ii) Fiabilité intra-observateur

Comme tous les clips vidéo ont été analysés par une seule personne (R.H.), pour tester la fiabilité intra-observateur, nous avons randomisé l'ordre des clips et remesuré la durée des gestes d'un clip sur neuf ($n = 102$ gestes sur 37 clips). Un test de coefficient de corrélation intraclasse (CCI) classe 3 avec $n = 1$ évaluateur [29] - a révélé une très grande concordance des mesures de la durée des gestes (CCI $\frac{1}{4}$ 0,975, $p < 0,0001$).

(iii) Définition des types de gestes et des jetons

Les linguistes font la distinction entre les types et les jetons [30,31]. Pour illustrer cela, considérons le vers du poème de Gertrude Stein "Sacred Emily" [31] : "Rose is a rose is a rose is a rose is a rose". Ce vers comprend 10 mots et trois types de mots différents. Les *types* sont les trois types de mots : *rose*, *est*, *a*. Les *jetons* représentent le nombre total de mots : 10 mots. Dans les recherches liées à la compression, les types sont utilisés pour tester la loi de Zipf sur les abréviations et pour calculer la durée moyenne, notée L [3,5,13,15]. Les jetons sont utilisés pour tester la loi de Menzerath et pour calculer la durée totale des jetons, notée M [4]. Nous avons donc pris en compte les types de gestes pour tester la loi de Zipf sur les abréviations et L , et les jetons de gestes pour tester la loi de Menzerath et M . Les types de gestes ont été définis comme des gestes qui se répètent dans la même forme de mouvement et qui sont utilisés seuls ou en séquence [32]. Nous avons considéré que les gestes simples représentaient une séquence de longueur 1, comme suit

Les séquences de gestes plus longues ont été définies comme deux gestes distincts ou plus, séparés par moins d'une seconde [25].

(d) Analyse

(i) Les types de gestes de jeu des chimpanzés suivent-ils la loi de Zipf sur les abréviations ?

Nous avons utilisé la corrélation de rang de Spearman bilatérale (IBM SPSS v. 22.0) pour déterminer si la durée moyenne et la fréquence d'utilisation des types de gestes étaient négativement corrélées. La durée moyenne de chaque type de geste a été calculée comme suit : $d \propto D/f$, où D est la somme de toutes les durées d'un type particulier et f est la fréquence d'utilisation de ce type (c'est-à-dire le nombre de fois que le geste est apparu dans notre ensemble de données) [33].

L'émergence de modèles conformes à la loi de Zipf sur l'abréviation dans des analyses de corrélation telles que celles-ci pourrait être un artefact de l'utilisation des valeurs moyennes de la durée du signal (ici, le geste). Plus précisément, une corrélation négative entre deux variables, $d \propto D/f$ et une autre f , peut être inévitable, étant donné que d est défini comme un quotient impliquant f , car alors $d \approx 1/f$ [34]. Cette explication peut être rejetée s'il peut être démontré que D et f sont significativement corrélés [33]. Pour toutes les analyses liées à la loi de Zipf sur les abréviations, nous avons donc testé l'existence d'une telle relation entre D et f , en utilisant la corrélation de rang de Spearman.

En plus de tester la loi de Zipf sur l'abréviation dans l'ensemble du répertoire gestuel, nous avons effectué d'autres analyses pour vérifier la présence de signes conformes à la loi dans des sous-ensembles spécifiques du répertoire. Ces analyses ont été effectuées car des études antérieures et des arguments théoriques indiquent que cette loi peut être trouvée dans des parties d'un répertoire de signaux même si elle n'est pas révélée par l'analyse de l'ensemble du répertoire [13]. Par exemple, chez une série d'espèces de chauves-souris, des modèles conformes à la loi n'apparaissent que lorsqu'un sous-ensemble spécifique du répertoire vocal - les appels sociaux - est pris en compte [14]. Chez les ouistitis communs, la loi n'a pas été trouvée dans les analyses de l'ensemble du répertoire vocal [35] mais a été trouvée par la suite dans un sous-ensemble du répertoire caractérisé par une faible durée totale (c'est-à-dire des appels avec un faible D) [13]. En plus de ces études empiriques, des arguments théoriques suggèrent que les modèles compatibles avec la loi de Zipf sur l'abréviation peuvent ne pas émerger si la pression pour la compression est compensée par d'autres pressions, par exemple la nécessité de maximiser le succès de la transmission et/ou d'atteindre des récepteurs éloignés, ce qui devrait entraîner une augmentation de la magnitude du signal [3] ; de telles pressions peuvent s'appliquer à certains signaux du répertoire, mais pas à d'autres.

Nous avons donc procédé à quatre analyses supplémentaires, en nous appuyant sur

La première a divisé le répertoire sur la base des valeurs D , en suivant l'approche générale de [13] et en se basant sur leurs conclusions que la loi de Zipf peut émerger dans des sous-ensembles du répertoire gestuel du jeu du chimpanzé. Le premier a divisé le répertoire en fonction des valeurs de D , en suivant l'approche générale de [13] et en se basant sur leurs conclusions selon lesquelles la loi peut émerger dans des sous-ensembles de répertoire à faible D mais pas à fort D . Le second a divisé le répertoire en fonction de la fréquence des abréviations. Le deuxième a divisé le répertoire en fonction de la fréquence d'utilisation des types de gestes, f , car la fréquence - ou un classement par fréquence - est un prédicteur fondamental de la longueur dans le contexte du codage optimal selon la théorie de l'information standard [22,36], tandis que dans les systèmes de communication naturels, la compression peut agir différemment sur les signaux, en fonction de la fréquence (ou non) à laquelle ils sont produits. La troisième analyse a divisé le répertoire en fonction de la durée moyenne des types de gestes, d , en raison de l'association entre f et d dans le contexte du codage optimal et aussi par souci d'exhaustivité, puisque $D \propto f d$. L'analyse finale a divisé le répertoire en fonction de la nature de la production des types de gestes - simples mouvements des membres et de la tête, connus sous le nom de "gestes manuels", ou mouvements impliquant l'ensemble du corps, connus sous le nom de "signaux du corps entier" [37] - car il a été proposé que les signaux de plus grande amplitude (comme c'est le cas pour les signaux du corps entier) puissent être moins susceptibles de révéler des schémas compatibles avec la compression [3].

Test de la loi de Zipf sur l'abréviation dans des sous-ensembles du répertoire en fonction des valeurs de D . Pour ces analyses, nous avons adapté la méthodologie de

[13]. Dans cette étude, la disposition des signaux par ordre de grandeur de D a révélé un point de rupture évident, délimitant une division entre un groupe "à fort D " et un groupe "à faible D ". Aucun point de rupture évident n'a été observé dans nos données (matériel électronique supplémentaire, S2), de sorte que nous n'avons pas pu effectuer une analyse similaire à celle de [13]. Nous avons donc adopté une autre approche, en commençant par classer les types de gestes par ordre croissant de D , puis en créant des sous-ensembles à partir de la valeur la plus faible de D jusqu'à la valeur la plus élevée de D , ou la procédure inverse (c'est-à-dire de la valeur la plus élevée à la valeur la plus faible de D). Ainsi, par exemple, les sous-ensembles commençant par la valeur D la plus basse contiennent : (i) le type de geste ayant la valeur D la plus basse, (ii) les deux types de geste ayant la valeur D la plus basse et la deuxième valeur D la plus basse, (iii) les 58 types de geste ayant la valeur D la plus basse, (iv) les 58 types de geste ayant la valeur D la plus basse,

2e plus bas 58e plus bas D (c.-à-d. tous les types de gestes). Pour tous les sous-ensembles

avec $n = 4$, nous avons utilisé la corrélation de rang de Spearman pour explorer la relation entre d et f .

Enfin, nous avons cherché à savoir si le modèle de résultats produit par un tel partitionnement fournissait des preuves de compression, ou s'il s'agissait plutôt d'un artefact du tri par D , en utilisant des tests de permutation mis en œuvre dans R (v. 3.2.3) (pour la justification, la méthode et le code R ; voir le matériel électronique supplémentaire, S3).

Test de la loi de Zipf sur l'abréviation dans les sous-ensembles du répertoire basés sur f et sur d . Nous avons suivi la méthodologie décrite ci-dessus pour D , afin de créer et d'analyser les sous-ensembles basés sur f et sur d . Comme précédemment, nous avons utilisé des tests de permutation pour vérifier si le schéma des résultats produits par les partitionnements fournissait des preuves de la pression, ou s'il pouvait être artéfactuel.

Test de la loi d'abréviation de Zipf dans des sous-ensembles du répertoire en fonction de la nature des gestes. Nous avons utilisé la corrélation de rang de Spearman pour tester la loi d'abréviation de Zipf dans les gestes manuels ($n = 44$ types ; matériel supplémentaire électronique, S4) et les signaux du corps entier ($n = 14$ types ; matériel supplémentaire électronique, S4).

(ii) La durée moyenne des types de gestes ludiques des chimpanzés est-elle significativement faible ?

Nous avons d'abord calculé la durée moyenne (L) des types de gestes, telle que définie dans l'équation (2.1) (d'après [3]), où n est le nombre d'éléments dans le répertoire, p_i est la fréquence normalisée du i ème élément le plus probable et e_i est la magnitude de cet élément [3]. La fréquence normalisée d'un type de geste a été estimée en divisant sa fréquence par la fréquence totale de tous les types de gestes. La magnitude d'un type de geste a été estimée par sa durée moyenne (s).

$$L = \sum_{i=1}^n p_i e_i. \tag{2.1}$$

Nous avons ensuite utilisé un test de permutation exécuté dans R (pour le code R, voir le matériel électronique supplémentaire, S5A) pour vérifier si L était significativement petit [38]. Un contrôle de L (L') a été défini sur la fonction de permutation $p(i)$, comme indiqué dans l'équation (2.2) [3]. La valeur p gauche a été calculée par Q_L/Q , Q_L étant le nombre de permutations uniformément aléatoires où $L' \leq L$, et Q le nombre total de permutations ($\approx 10^5$). La valeur p droite a été calculée par Q_R/Q , Q_R étant le nombre de permutations aléatoires où $L' \geq L$, et Q le nombre total de permutations ($\approx 10^5$).⁵

$$L' = p \sum_{i=1}^n e_{i(p(i))} \tag{2.2}$$

(iii) Les séquences de gestes de jeu des chimpanzés suivent-elles la loi de Menzerath ?

Nous avons utilisé la corrélation de rang de Spearman pour déterminer si la taille de la séquence (nombre de gestes) et la durée moyenne du geste étaient négativement corrélées. Il n'est pas nécessaire de savoir si un seul geste est corrélé à la durée moyenne du geste.

Tableau 1. Fréquence, n , des séquences de gestes en fonction de leur taille (nombre de gestes dans la séquence).

| taille de la séquence | n |
|-----------------------|-----|
| 1 | 873 |
| 2 | 267 |
| 3 | 93 |
| 4 | 42 |
| 5 | 17 |
| 6 | 10 |
| 7 | 4 |
| 8 | 1 |
| 9 | 3 |
| 14 | 1 |
| 16 | 1 |
| 45 | 1 |

Les signaux doivent être considérés comme des séquences (c'est-à-dire de taille 1). Les analyses ont donc été effectuées à la fois pour l'ensemble des données (c'est-à-dire les séquences de toutes tailles, y compris les gestes uniques) et pour un ensemble de données excluant les gestes uniques (c'est-à-dire les séquences de deux gestes ou plus).

Dans le contexte de la loi de Menzerath sur les gestes des chimpanzés, D est défini comme la durée totale des gestes dans une séquence (à l'exclusion des durées des intervalles entre les gestes consécutifs), d comme la durée moyenne des gestes et n comme le nombre de gestes dans cette séquence. La loi de Menzerath s'applique s'il existe une corrélation négative significative entre $d \propto D/n$ et n . Il a été avancé que des modèles conformes à la loi de Menzerath pourraient émerger comme une conséquence inévitable de l'exploration de la relation entre des variables telles que n et $d \propto D/n$, car d s'adapterait automatiquement à n comme $d \approx 1/n$ [34]. Cependant, une analyse mathématique rigoureuse a montré que cela ne peut se produire que dans une condition très spéciale, à savoir lorsque D est la moyenne indépendante de n , une propriété qui peut être testée avec un simple test de corrélation entre D et n [39]. Pour exclure cette explication simpliste de la loi de Menzerath, une analyse plus poussée a été effectuée [39,40], en suivant les méthodes utilisées pour explorer la robustesse des résultats relatifs à la loi de Menzerath dans les génomes [41]. Pour vérifier si la loi de Menzerath dans les séquences gestuelles des chimpanzés est une conséquence inévitable d'une mise à l'échelle triviale, nous avons utilisé la corrélation de rang de Spearman pour tester la relation entre D et n ; une relation négative significative exclut l'explication triviale.

(iv) La somme totale attendue de la durée des gestes de chaque séquence est-elle significativement faible ?

La durée totale d'une collection de séquences est définie comme suit

$$M = \sum_{i=1}^{\infty} D_i \quad (2.3)$$

où D_i est la durée totale de la i ème séquence.

À son tour,

$$D_i = \sum_{j=1}^{\infty} l_{ij} \quad (2.4)$$

où n_i est le nombre d'éléments de la i ème séquence et l_{ij} est la durée du j ème élément de la i ème séquence. En définissant la durée moyenne de la i ème séquence comme $\langle l_{ij} \rangle_i = D_i/n_i$, on peut définir M comme suit

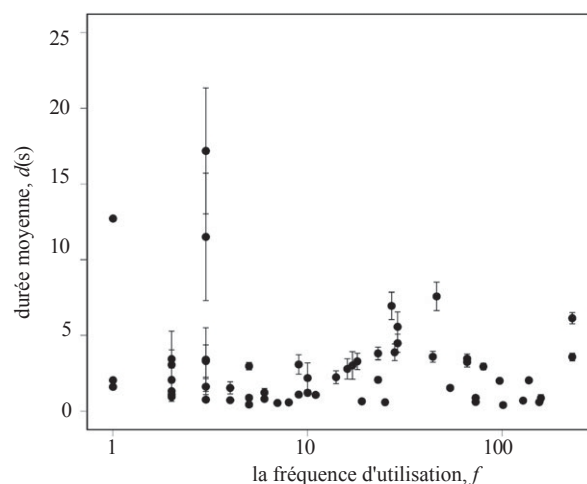


Figure 1. Relation entre la durée moyenne (d) et la fréquence d'utilisation (f) des types de gestes. L'axe des x est représenté en échelle logarithmique. Les moustaches indiquent

L'absence de moustaches indique soit une faible variation des durées au sein d'un type de geste, soit qu'un type de geste n'a été utilisé qu'une seule fois.

exprimée comme suit

$$M = \sum_{i=1}^{\infty} n_i \langle l_{ij} \rangle_i \quad (2.5)$$

Pour vérifier si la somme totale de la durée des gestes de chaque séquence est significativement faible, nous avons calculé M en suivant l'équation (2.5). Le calcul de M est défini sur une somme de jetons, chaque occurrence d'une séquence étant considérée comme un jeton individuel. Nous avons utilisé un test de permutation similaire à celui utilisé pour tester la signification de L , exécuté en R (pour le code R, voir le matériel complémentaire électronique, S5B), afin de vérifier si M était significativement petit [3]. n_i joue le rôle de p_i et $\langle l_{ij} \rangle_i$ joue le rôle de e_i dans le test. Ainsi, n_i et $\langle l_{ij} \rangle_i$ restent constants pendant le test.

3. Résultats

Les durées de 2137 gestes de jeu ont été mesurées; elles comprenaient 58 types de gestes, effectués par 48 chimpanzés individuels. Sur ces 2137 gestes, 873 étaient des gestes isolés et les 1264 restants des séquences allant de 2 à 45 gestes (tableau 1). Les bébés ont produit 492 (23,02%) gestes, les jeunes 940 (23,02%) (43,99%), les subadultes 638 (29,85%) et les adultes 67 (3,14%) (matériel complémentaire électronique, S4).

(a) Les gestes de jeu suivent-ils la loi de Zipf sur les abréviations ?

(i) Test de la loi de Zipf sur l'abréviation dans le répertoire général

Il n'y avait pas de corrélation significative entre la durée moyenne (d) et la fréquence d'utilisation (f) des types de gestes ($r_s \propto 20,005$, $n \propto 58$, $p \propto 0,97$, figure 1). La durée moyenne des types de gestes (L) était de 2,65 s, ce qui n'est pas significativement faible ($n \propto 58$, $p \propto 0,42$).

(ii) Test de la loi de Zipf sur l'abréviation dans des sous-ensembles du répertoire sur la base des valeurs de D

La loi de Zipf sur l'abréviation était prédominante dans les sous-ensembles de types de gestes à faible D (figure 2a; pour les résultats complets, voir le matériel électronique complémentaire, S6). Si l'on considère les sous-ensembles successifs de types de gestes générés de faible D à fort D (avec $n \geq 4$), une corrélation négative significative entre d et f a d'abord été observée dans le sous-ensemble comprenant les gestes avec les cinq plus faibles D (figure 2a), puis dans le sous-ensemble comprenant les cinq plus faibles D (figure 2b).

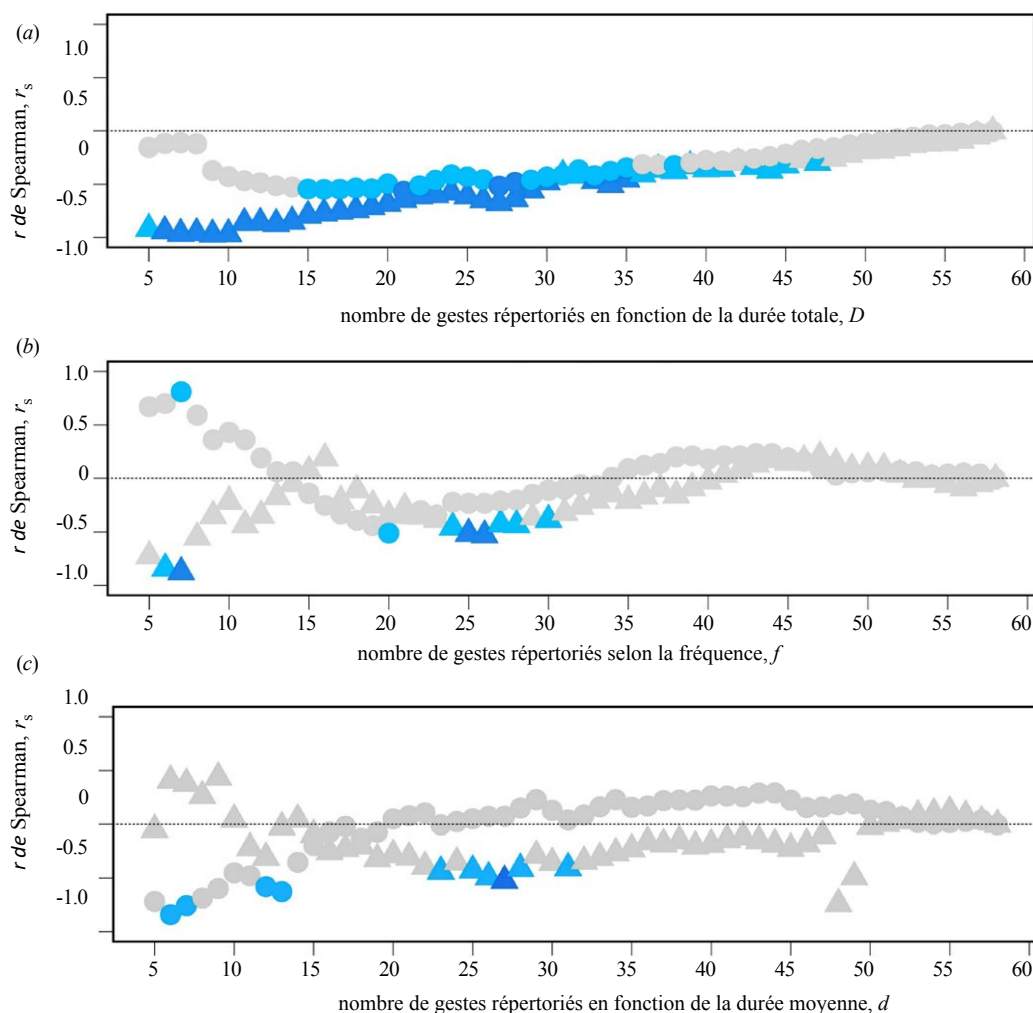


Figure 2. Coefficients de corrélation (corrélation de rang de Spearman) entre la durée moyenne (d) et la fréquence d'utilisation (f) des types de gestes, pour les sous-ensembles de types de gestes générés soit par l'inclusion incrémentale de types de gestes du plus petit au plus grand (triangles) ou du plus grand au plus petit (cercles) des valeurs de (a) D , (b) f et (c) d . Les symboles en gris indiquent $p < 0,05$, en bleu clair indiquent $0,05 < p < 0,01$ et en bleu foncé indiquent $p > 0,01$.

Au fur et à mesure que les types de gestes ayant des valeurs D plus élevées étaient ajoutés un par un, la corrélation entre d et f est restée significative jusqu'au sous-ensemble de types de gestes ayant les 41 valeurs D les plus faibles, après quoi les valeurs p ont fluctué autour de 0,05 jusqu'au sous-ensemble ayant les 48 valeurs D les plus faibles, à partir de quoi toutes les corrélations étaient non significatives. L était significativement faible pour tous les sous-ensembles de types de gestes générés de D faible à D élevé, jusqu'à celui des types de gestes avec les 55 valeurs de D les plus faibles (matériel électronique supplémentaire, S6). D et f étaient significativement corrélés - et donc l'accord avec la loi de Zipf sur l'abréviation ne semble pas être un artefact de l'analyse de la durée moyenne des gestes - pour le sous-ensemble contenant les types de gestes avec les neuf valeurs les plus faibles de D et pour tous les sous-ensembles plus importants (matériel supplémentaire électronique, S6). En outre, les tests par mutations ont fourni des preuves de compression dans un large éventail de sous-ensembles de types de gestes avec les valeurs les plus basses de D (matériel électronique supplémentaire, S3).

Le schéma des résultats dans les sous-ensembles générés dans la direction opposée - d'une valeur D élevée à une valeur D faible - était quelque peu différent (figure 2a ; pour les résultats complets, voir le matériel électronique supplémentaire, S6). Une corrélation négative significative entre d et f n'a pas été observée avant le sous-ensemble contenant les 15 types de gestes ayant les valeurs D les plus élevées ; la corrélation est restée significative jusqu'au sous-ensemble contenant les 35 types de gestes ayant la valeur D la plus élevée, puis - à l'exception du sous-ensemble contenant les 38 types de gestes ayant la valeur D la plus élevée - jusqu'au sous-ensemble contenant les 35 types de gestes ayant la valeur D la plus élevée.

était non significatif dans tous les autres sous-ensembles, de plus en plus grands. L était significativement faible dans le sous-ensemble contenant les 21 types de gestes ayant la valeur D la plus élevée jusqu'au sous-ensemble contenant les 38 types de gestes ayant la valeur D la plus élevée (matériel électronique supplémentaire, S6). Alors que D et f étaient significativement corrélés pour les sous-ensembles contenant les gestes avec les 10, 11, 13, 14 et 20 valeurs les plus élevées de D et pour tous les sous-ensembles plus grands (matériel supplémentaire électronique, S6), les tests de permutation n'ont pas fourni de preuve de compression dans les sous-ensembles de types de gestes avec des valeurs élevées de D , ce qui indique que les corrélations significatives dans ces sous-ensembles sont un artefact du processus de tri (matériel supplémentaire électronique, S3).

(iii) Test de la loi de Zipf sur l'abréviation dans des sous-ensembles du répertoire basés sur les valeurs de f

Lorsque les gestes ont été regroupés par ordre de f , des relations négatives significatives entre d et f n'ont été trouvées que dans un petit nombre de sous-ensembles et les tests de permutation n'ont pas fourni de preuves de compression (figure 2b ; pour les résultats complets, et les calculs de L et les corrélations entre D et f , voir le matériel électronique supplémentaire, S7 ; pour les résultats des tests de permutation, voir le matériel électronique supplémentaire, S3).

(iv) Test de la loi de Zipf sur l'abréviation dans des sous-ensembles du répertoire basés sur les valeurs de d

L'analyse des sous-ensembles de gestes regroupés en fonction de d n'a révélé que quelques relations négatives significatives entre d et f . Les tests de permutation ont mis en évidence une compression dans une fourchette étroite des sous-ensembles de types de gestes présentant les valeurs de d les plus élevées, mais pas ailleurs (figure 2c ; pour les résultats complets, et pour les calculs de L et les corrélations entre D et f , voir le matériel électronique supplémentaire, S8 ; pour les résultats des tests de permutation, voir le matériel électronique supplémentaire, S3).

(v) Test de la loi de Zipf sur l'abréviation dans des sous-ensembles du répertoire basés sur la nature des gestes

L'analyse des gestes manuels n'a révélé aucune relation entre d et f (r_s $\frac{1}{4}$ 20.125, n $\frac{1}{4}$ 44, p $\frac{1}{4}$ 0.419) et L n'était pas significativement petit (2.09 s, p $\frac{1}{4}$ 0.148). De façon inattendue, les signaux du corps entier ont montré une relation positive significative entre d et f (r_s $\frac{1}{4}$ 0.746, n $\frac{1}{4}$ 14, p $\frac{1}{4}$ 0,002) - le modèle opposé à celui prédit par la loi d'abréviation de Zipf - et L était significativement grand (5,29 s, p $\frac{1}{4}$ 0,0001).

(b) Les séquences de gestes de jeu des chimpanzés suivent-elles la loi de Menzerath ?

Il existe une corrélation négative significative entre la séquence (n_i) et la durée moyenne des gestes constitutifs ($(l_{ij})_i$), à la fois en incluant les gestes uniques (r_s $\frac{1}{4}$ 20.077, n $\frac{1}{4}$ 1313, p $\frac{1}{4}$ 0.006-figure 3) et en excluant les gestes uniques (r_s $\frac{1}{4}$ 20.156, n $\frac{1}{4}$ 440, p $\frac{1}{4}$ 0.001). Ces relations sont restées inchangées significative après avoir retiré le point de données aberrant - une séquence comprenant 45 gestes (y compris les gestes uniques : r_s $\frac{1}{4}$ 20,074, n $\frac{1}{4}$ 1312, p $\frac{1}{4}$ 0,007 ; à l'exclusion des gestes uniques : r_s $\frac{1}{4}$ 20,149, n $\frac{1}{4}$ 439, p $\frac{1}{4}$ 0,002).

Il existe une corrélation positive significative entre la taille de la séquence (n_i) et la durée totale des gestes constitutifs (D_i) (y compris les gestes *uniques*- r_s $\frac{1}{4}$ 0,403, n $\frac{1}{4}$ 1313 et p $\frac{1}{4}$ 0,0001 ; à l'exclusion des gestes *uniques*- r_s $\frac{1}{4}$ 0,209, n $\frac{1}{4}$ 440 et p $\frac{1}{4}$ 0,0001), ce qui confirme que la loi de Menzerath n'est pas un artefact d'une mise à l'échelle inévitable et triviale.

La somme totale de la durée des gestes de chaque séquence, M , était de 5653,82 s dans l'ensemble de données complet et de 3050,06 s dans l'ensemble de données excluant les gestes uniques ; les deux valeurs de M étaient significativement petites (incluant les gestes uniques : n $\frac{1}{4}$ 1313, p $\frac{1}{4}$ 0,0001 ; excluant les gestes uniques : n $\frac{1}{4}$ 440, p $\frac{1}{4}$ 0,0001).

4. Discussion

Nous avons recherché des preuves de compression dans la communication gestuelle ludique des chimpanzés, d'abord en examinant si les types de gestes et les séquences de gestes suivent des lois linguistiques qui reflètent ce principe, et ensuite en vérifiant si les mesures de la longueur moyenne des codes des types et des séquences sont significativement petites. Il a d'abord été constaté que les types de gestes individuels ne suivaient pas la loi d'abréviation de Zipf (qui prédit une relation négative entre la longueur du signal et la fréquence d'utilisation) ; cependant, des analyses ultérieures de sous-ensembles spécifiques du répertoire gestuel global ont révélé une forte concordance avec cette loi, ainsi que des preuves que la longueur moyenne du code - ici, la durée du geste - était significativement petite. De manière inattendue, des modèles opposés à la loi ont été trouvés dans un sous-ensemble de gestes, les signaux du corps entier.

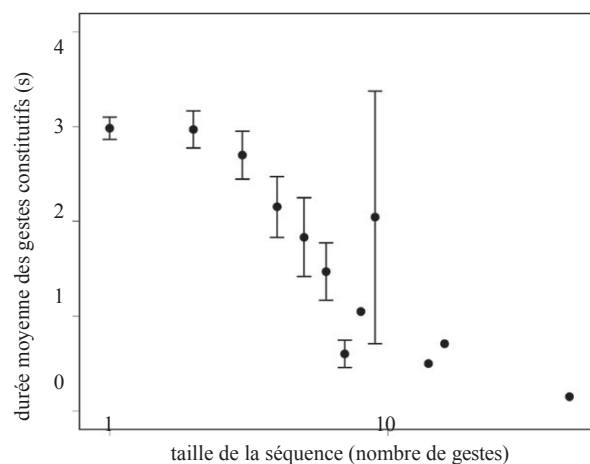


Figure 3. Relation entre la durée moyenne des gestes constitutifs et la taille de la séquence en termes de nombre de gestes dans les séquences de jeu. L'axe des x est représenté en échelle logarithmique. L'absence de moustaches indique qu'une séquence de cette taille n'a été utilisée qu'une seule fois.

Les séquences de gestes suivaient la loi de Menzerath (selon laquelle les séquences plus longues sont composées de constituants plus courts), et encore une fois la longueur moyenne du code - ici la somme totale de la durée des gestes - était significativement petite. Ces résultats indiquent que la compression a façonné la communication gestuelle ludique des chimpanzés à deux niveaux d'organisation : le modèle d'utilisation des types de gestes individuels et la construction de séquences de gestes. Nos résultats étendent la preuve de la compression dans la communication animale pour la première fois au mode gestuel de signalisation ; en conjonction avec les résultats des études du comportement non vocal chez les dauphins [15], une gamme de systèmes vocaux animaux [3 - 5,14,18], la parole humaine [42], les textes écrits [9,10] et le langage des signes [12], ce travail apporte un soutien supplémentaire à l'hypothèse que la compression est un principe général à la base de diverses formes et modalités de la communication.

Une telle hypothèse est étayée par de fortes prédictions de la théorie de l'information en relation avec trois lois linguistiques. En ce qui concerne la loi de Zipf sur les abréviations, ces prédictions ont en commun le fait que le codage optimal de l'information (L minimum) implique que la corrélation entre la fréquence relative d'un type, p , et sa longueur l , ne peut pas être positive [22]. La théorie standard de l'information est capable de prédire la relation réelle entre la longueur et la fréquence dans le cas d'un système entièrement optimisé. Dans le cas d'un codage optimal, uniquement déchiffirable, l devrait avoisiner $2\log p$ [36]. Dans le cas d'un codage optimal non singulier, la longueur d'un type de rang de fréquence i (le type le plus fréquent a le rang 1) devrait être proche de $\log i$ [22]. Ces arguments ont été étendus pour prédire la loi de Menzerath à partir d'un codage optimal (minimum M) [4]. Enfin, la loi de Zipf, bien étudiée et omniprésente pour les fréquences de mots, peut également être une conséquence de la compression [43]. Ce mécanisme puissant et abstrait est étayé par l'omniprésence de la loi d'abréviation dans le langage humain, indépendamment de la modalité (parlé ou signé) [12,42] ou du système d'écriture (basé sur les caractères ou sur les lettres) [6,9,10].

Les résultats de nos analyses de la loi de Zipf sur les abréviations réitèrent un point clé soulevé par des études antérieures [13,14], à savoir que l'exploration des lois linguistiques dans les systèmes non humains peut nécessiter l'étude de modèles à des niveaux inférieurs au répertoire complet des signaux. Dans l'ensemble, les types de gestes de jeu individuels des chimpanzés n'étaient pas conformes au modèle

Cependant, une très forte concordance a été observée dans des sous-ensembles du répertoire, en particulier ceux pour lesquels D , le produit de la durée moyenne (d) et de la fréquence d'utilisation (f), était faible. En revanche, les analyses des sous-ensembles basées sur d et f ont révélé une faible concordance avec la loi de Zipf sur les abréviations. D peut être considéré comme une fonction de "coût total", et il peut sembler contre-intuitif que ce soit dans les gestes à faible coût total que la compression soit la plus répandue ; des économies plus importantes en termes d'efficacité de codage pourraient, en principe, être réalisées dans les gestes à coût total élevé. Cependant, il est possible que les gestes à faible D soient à faible D précisément à cause de la compression ; ce principe peut avoir agi pour améliorer l'efficacité du codage non seulement en alignant la fréquence d'utilisation et la durée de ces gestes, mais aussi en réduisant ces deux mesures (et donc leur produit) dans l'ensemble.

D'autre part, il peut y avoir des raisons pour lesquelles, parmi d'autres gestes, on ne trouve pas de modèles compatibles avec la compression. Une possibilité est que la compression n'affecte pas ces gestes, contrairement à la proposition récente selon laquelle la compression est un principe universel qui sous-tend non seulement le comportement animal [3], mais aussi les systèmes d'information biologiques au sens large [4]. En effet, la pression pour l'efficacité peut être réduite dans le contexte que nous avons exploré - le jeu social - puisque ce comportement est associé à un excès de temps et/ou d'énergie [44]. Toutefois, l'absence d'accord avec la loi de Zipf sur l'abréviation n'exclut pas que la compression agisse dans un système. Les principes universels ne produisent pas nécessairement des modèles universels. Même lorsqu'un principe est valable, d'autres forces peuvent conduire à l'émergence de modèles qui sont, superficiellement, incompatibles avec ceux prédits par le principe seul [3] ; cette situation apparaît de manière récurrente dans les modèles d'optimisation de la communication [45,46]. Le défi consiste à identifier ces forces et à déterminer dans quelles circonstances elles l'emportent sur le principe sous-jacent [3].

L'compression est la minimisation de la redondance dans un système, et l'absence d'un modèle prédit par ce principe parmi un ensemble de gestes de jeu de chimpanzés (et le répertoire dans son ensemble) peut être due à l'ajout de redondance dans certaines parties du répertoire. La théorie du codage indique que l'ajout de redondance dans les signaux, par exemple en les allongeant, réduit le risque d'erreurs de transmission [36]. Dans notre système d'étude, de telles erreurs peuvent être coûteuses car le jeu social peut devenir brutal et peut, dans les cas extrêmes, conduire à des escalades agressives [47]. Les types de gestes régulateurs utilisés pendant le jeu (par exemple la *position de la tête*, le *balancement*, le *retournement*) ou qui signalent l'arrêt ou le changement du jeu (par exemple la *main sur*) peuvent donc devoir être utilisés de manière très explicite pour assurer la poursuite d'un jeu pacifique [32] ; en particulier, ces types de gestes ont tendance à être caractérisés par une D élevée. Le coût de l'ajout de redondance à certains gestes (augmentation de leur durée) peut donc être compensé par le coût de l'agression résultant d'un signal mal interprété.

Le regroupement des gestes en fonction de leur fréquence d'apparition, f , n'a pas produit de modèles clairs de concordance avec la loi de Zipf sur les abréviations. Ce résultat est surprenant, car un certain nombre de résultats de la théorie standard de l'information associent la fréquence d'un type, f , à sa longueur dans le contexte d'un codage optimal. Par exemple, la longueur d'un type dont la fréquence relative est p devrait avoisiner $2\log p$ dans le cas d'un codage optimal uniquement déchiffirable [36] ; et la longueur d'un type de rang de fréquence i (le type le plus fréquent a le rang 1) devrait avoisiner $\log i$ dans le cas d'un codage optimal non singulier [22]. Ces résultats indiquent que la fréquence, ou un ordre induit par la fréquence, est fondamentale pour le codage non singulier.

théorie de l'information standard et, par conséquent, on peut s'attendre à ce que cette 7 de la communication animale.

Cependant, nos résultats suggèrent que D capture la pression d'optimisation beaucoup mieux que f dans ce système biologique du monde réel. Une explication possible intrigante est que les conclusions de la théorie standard de l'information ne peuvent pas être complètement extrapolées à de tels systèmes, par exemple parce que les hypothèses de la théorie peuvent ne pas être valides. La théorie de l'information standard fournit une approche à sens unique du codage optimal : elle fournit les longueurs minimales de la chaîne de chaque type, compte tenu de la probabilité des types. Ainsi, la longueur d'un type dépend de sa fréquence, et non l'inverse. Cependant, les fréquences des types varient dans les systèmes de communication naturels et, par conséquent, dans ces systèmes, il peut y avoir des pressions reflétant une solution bidirectionnelle au codage optimal : la fréquence du type peut influencer la longueur de sa chaîne (comme dans la théorie de l'information standard) et *vice versa* - la longueur de la chaîne d'un type peut influencer sa fréquence. Dans un système d'optimisation à double sens, la sélection naturelle opérerait sur le produit de la fréquence et de la durée, et non sur la durée ou la fréquence seule.

Les analyses des gestes regroupés par la durée moyenne, d , ont également révélé peu d'accord avec la loi de Zipf sur l'abréviation. La faible performance de d dans la détection de l'accord avec cette loi n'est pas surprenante puisque la théorie de l'information prédit une forte corrélation entre d et f (ou entre d et le rang de fréquence), et si f n'a pas réussi à partitionner le répertoire d'une manière qui révèle l'accord avec l'abréviation de la loi de Zipf, la même chose devrait s'appliquer à d . Nos résultats pour les sous-ensembles regroupés par d , en conjonction avec ceux regroupés par f , indiquent que ce n'est pas parmi les appels qui sont en moyenne courts, ou ceux qui sont rarement donnés, que la compression est la plus évidente, mais plutôt parmi les appels où les deux choses sont le cas (le produit de f et de d est D).

Les analyses de la loi de Zipf sur l'abréviation dans les gestes manuels et dans les signaux du corps entier n'ont révélé aucune preuve de la loi dans le premier cas, mais un modèle opposé à celui prédit par la loi dans le second. Alors que certaines études antérieures sur la communication animale n'ont pas confirmé cette loi [3], à notre connaissance, c'est la première fois qu'une relation positive significative a été trouvée dans la communication humaine ou non humaine entre la durée du signal et la fréquence d'utilisation. Ce résultat fournit des preuves irréfutables pour réfuter les propositions

[34] que les modèles compatibles avec les lois linguistiques sont inévitables, et donc que ces lois sont scientifiquement triviales. Un schéma contraire à celui prédit par la loi de Zipf sur l'abréviation peut provenir d'un certain nombre de sources : la redondance peut avoir été ajoutée en relation positive avec la fréquence d'utilisation (c'est-à-dire que les signaux du corps entier les plus courants comportent le plus grand degré de redondance) ; la compression peut agir en relation positive avec la rareté (c'est-à-dire que les signaux du corps entier les plus rarement utilisés sont les plus compressés) ; ou les deux pressions peuvent être à l'œuvre. Un facteur clé à prendre en compte en ce qui concerne les signaux du corps entier est que certains d'entre eux exigent qu'une posture soit maintenue en place pour être clairement identifiés comme un signal spécifique ; par exemple, un *lever de tête* n'est clairement qu'un *lever de tête*, et non la moitié d'un *saut périlleux* ou d'un autre mouvement, parce que le signaleur s'arrête dans la position inhabituelle de se tenir debout avec la tête entre les pieds et qu'il maintient cette position. Cette extension inévitable de certains signaux - potentiellement en conjonction avec une absence ou un relâchement des contraintes énergétiques [44] - peut expliquer l'association positive entre la durée du signal du corps entier et la fréquence d'utilisation.

Nous avons constaté que les séquences gestuelles de jeu des chimpanzés suivent la loi de Menzerath, une loi linguistique dérivée pour la première fois de l'article de la Convention de Genève sur les droits de l'homme.

Cette loi, qui a été étudiée dans le cadre d'études sur le langage humain et dont il a été récemment démontré qu'elle s'appliquait également aux séquences vocales des géladas [4] et des chimpanzés [18], suggère que des principes comparables d'auto-organisation [48] sous-tendent ces différents systèmes de communication combinatoire. Cette loi n'avait jamais été étudiée dans la communication gestuelle chez l'homme ou chez d'autres espèces ; nos résultats apportent de nouvelles preuves d'un point commun important entre le langage humain et la communication gestuelle des primates, en ce qui concerne les modèles structurels de base qui sous-tendent la manière dont les signaux sont combinés en structures plus vastes. Dans les études de cette loi dans la communication vocale des primates [4,18], les contraintes liées à la respiration et les exigences énergétiques de la production vocale ont été impliquées comme des moteurs importants de la relation négative entre le nombre d'appels dans une séquence et leur durée moyenne. Les séquences gestuelles ne sont pas contraintes par des schémas respiratoires, comme c'est le cas pour les séquences vocales. Les contraintes énergétiques, associées à l'augmentation de l'activité musculaire impliquée dans la production de gestes, et en particulier les épisodes prolongés de gestes [49], peuvent sous-tendre l'émergence de la loi de Menzerath dans ce système.

Notre travail s'ajoute à une littérature de plus en plus abondante dans laquelle les lois statistiques dérivées des études sur le langage humain s'avèrent valables pour les systèmes non humains [3 - 5,14,15,18]. L'identification de propriétés communes partagées entre le langage et d'autres systèmes naturels, et l'examen des fondements mathématiques de ces propriétés, permettent non seulement de mieux comprendre les principes fondamentaux de l'organisation naturelle [3], mais offrent également une occasion importante d'explorer l'histoire évolutive des modèles linguistiques universels [4]. De nombreuses lois linguistiques restent

Nous espérons que notre travail **8** encouragera de telles recherches dans divers systèmes d'information biologiques.

Éthique. La collecte des données a été autorisée par le comité d'éthique et de bien-être des animaux de l'université de St Andrews.

Accessibilité des données. Les codes R à l'appui de cet article se trouvent dans le matériel électronique supplémentaire. Les ensembles de données sont publiés à l'adresse suivante : <https://doi.org/10.6084/m9.figshare.5970823>.

Contributions des auteurs. R.H. et S.S. ont conçu l'étude ; R.H., S.S. et R.F.C. a conçu l'étude ; C.H. a fourni les données brutes (enregistrements vidéo et classification des gestes) ; R.H. et R.F.C. ont analysé les données ; R.H. et S.S. ont rédigé l'article, révisé par R.F.C. et C.H. Tous les auteurs ont donné leur accord final pour la publication.

Intérêts concurrents. Nous déclarons ne pas avoir d'intérêts concurrents.

Financement. R.F.C. est financé par les subventions 2014SGR 890 (MACDA) de l'AGAU (Generalitat de Catalunya) et la subvention Gestion et analyse de données complexes (TIN2017-89244-R) du MINECO (Ministerio Econom'ia Industria y Competitividad). Le travail sur le terrain de C.H. a été généreusement soutenu par des subventions de la Fondation Wenner-Gren et du Russell Trust. S.S. remercie Santander pour une bourse de recherche et de voyage utilisée pour travailler sur cette étude. R.H. remercie Ko' Iner Gymnasial- und Stiftungsfonds pour son soutien financier.

Remerciements. Nous remercions le Dr Peter Shaw pour ses conseils statistiques, le Dr Emilie Genty pour ses conseils sur la classification des signaux du corps entier et Nous remercions également le Conseil national ougandais pour la science et la technologie et l'Autorité ougandaise de la faune pour nous avoir autorisés à collecter les données, ainsi que la Royal Zoological Society of Scotland et le personnel de la Budongo Conservation Field Station pour leur soutien inestimable. Nous remercions le Dr David Leavens pour avoir facilité cette collaboration, ainsi que deux évaluateurs anonymes pour leurs commentaires et réactions réfléchies. Nous remercions également Argimiro Arratia pour ses conseils sur R.

Références

- Koehler R, Altmann G, Piotrowski RG. 2005 *Quantitative linguistics : an international handbook*. New York, NY : Walter de Gruyter.
- Altmann EG, Gerlach M. 2016 *Statistical laws in linguistics*, pp. 7 - 26. Cham, Suisse : Springer.
- Ferrer-i-Cancho R, Hernánde-Ferna'ndez A, Lusseu D, Agoramoorthy G, Hsu MJ, Semple S. 2013 Compression as a universal principle of animal behavior. *Cogn. Sci.* **37**, 1565 - 1578. (doi:10.1111/cogs.12061)
- Gustison ML, Semple S, Ferrer-i-Cancho R, Bergman TJ. 2016 Gelada vocal sequences follow Menzerath's linguistic law. *Proc. Natl Acad. Sci. USA* **113**, E2750 - E2758. (doi:10.1073/pnas.1522072113)
- Semple S, Hsu MJ, Agoramoorthy G. 2010 Efficacité du codage dans la communication vocale des macaques. *Biol. Lett.* **6**, 469 - 471. (doi:10.1098/rsbl.2009.1062)
- Zipf G. 1936 *The psycho-biology of language : an introduction to dynamic philology*. Londres, Royaume-Uni : George Routledge and Sons Ltd.
- Zipf G. 1949 *Human behaviour and the principle of least effort*. Cambridge, MA : Addison Wesley.
- Bentz C, Ferrer-i-Cancho R. 2015 La loi de Zipf sur les abréviations en tant qu'universalité linguistique. Voir <https://publikationen.uni-tuebingen.de/xmlui/handle/10900/68558>.
- Wang Y, Chen X. 2015 Structural complexity of simplified Chinese characters. In *Recent contributions to quantitative linguistics* (eds A Tuzzi, M Bene'sova', J Macutek), pp. 229 - 240. Boston, MA : Walter de Gruyter.
- Sanada H. 2008 *Investigations in Japanese historical lexicology*. Go'ttingen, Allemagne : Peust & Gutschmidt.
- Hernánde-Ferna'ndez A, Casas B, Ferrer-i-Cancho R, Baixerias J. 2016 *Testing the robustness of laws of polysemy and brevity versus frequency*, pp. 19 - 29. Cham, Suisse : Springer.
- Bo'stall C, Ho'rborg T, Ostling R. 2016 Distribution et durée des signes et des parties du discours dans la langue des signes suédoise. *Sign Lang. Linguist.* **19**, 143 - 196. (doi:10.1075/sll.19.2.01bor)
- Ferrer-i-Cancho R, Hernánde-Ferna'ndez A. 2013 L'échec de la loi de la brièveté chez deux primates du Nouveau Monde. *Statistical caveats. Glottotheory* **4**, 44 - 55. (doi:10.1524/glot.2013.0004)
- Luo B, Jiang T, Liu Y, Wang J, Lin A, Wei X, Feng J. 2013 Brevity is prevalent in bat short-range communication. *J. Comp. Physiol. A* **199**, 325 - 333. (doi:10.1007/s00359-013-0793-y)
- Ferrer-i-Cancho R, Lusseu D. 2009 Efficient coding in dolphin surface behavioral patterns. *Complexité* **14**, 23 - 25. (doi:10.1002/cplx.20266)
- Altmann G. 1980 Prolégomènes à la loi de Menzerath. *Glottometrika* **2**, 1 - 10.
- Teupenhayn R, Altmann G. 1984 Clause length and Menzerath's law. *Glottometrika* **6**, 127 - 138.
- Fedurek P, Zuberbu'bler K, Semple S. 2017 Trade-offs in the production of animal vocal sequences : insights from the structure of wild chimpanzee pant hoots. *Front. Zool.* **14**, 50. (doi:10.1186/s12983-017-0235-8)
- Ferrer-i-Cancho R, Forns N. 2010 The self-organisation of genomes. *Complexité* **15**, 34 - 36. (doi:10.1002/cplx.20296)
- Li W. 2012 Menzerath's law at the gene-exon level in the human genome. *Complexité* **17**, 49 - 53. (doi:10.1002/cplx.20398)
- Shahzad K, Mittenenthal JE, Caetano-Anolle's G. 2015 L'organisation des domaines dans les protéines obéit à la loi du langage de Menzerath-Altmann. *BMC Syst. Biol.* **9**, 44 (doi:10.1186/s12918-015-0192-9).
- Ferrer-i-Cancho R, Bentz C, Sequin C. 2015 Compression et origines de la loi de Zipf sur les abréviations. Voir <https://arxiv.org/abs/1504.04884>.
- Arbib MA, Liebal K, Pika S. 2008 Primate vocalization, gesture, and the evolution of human language. *Curr. Anthropol.* **49**, 1053 - 1076. (doi:10.1086/593015)
- Call J, Tomasello M. 2007 *The gestural communication of apes and monkeys*. Mahwah, NJ : Lawrence Erlbaum Associates.
- Hobaiter C, Byrne RW. 2011 Gestes en série chez les chimpanzés sauvages : leur nature et leur fonction de communication. *Anim. Cogn.* **14**, 827 - 838. (doi:10.1007/s10071-011-0416-3)

26. Altmann J. 1974 Observational study of behavior : sampling methods. *Behaviour* **49**, 227 - 266. (doi:10.1163/156853974X00534)
27. Hayaki H. 1985 Social play of juvenile and adolescent chimpanzees in the Mahale Mountains National Park, Tanzania. *Primates* **26**, 343 - 360. (doi:10.1007/BF02382452)
28. Hobaiter C, Byrne RW. 2011 Le répertoire gestuel du chimpanzé sauvage. *Anim. Cogn.* **14**, 745 - 767 (doi:10.1007/s10071-011-0409-2).
29. Landers RN. 2015 Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *The Winnower* **2**, e143518. (doi:10.15200/winn.143518.81744)
30. Brainerd B. 1972 On the relation between types and tokens in literary text. *J. Appl. Probab.* **9**, 507 - 518. (doi:10.1017/S002190020003583X)
31. Wetzel L. 2011 Types and tokens. *The Stanford encyclopedia of philosophy* (ed. E Zalta). Voir <http://plato.stanford.edu/archives/spr2011/entries/types-tokens>.
32. Hobaiter C, Byrne RW. 2014 La signification des gestes des chimpanzés. *Curr. Biol.* **24**, 1596 - 1600. (doi:10.1016/j.cub.2014.05.066)
33. Semple S, Hsu MJ, Agoramoorthy G, Ferrer-i-Cancho R. 2013 La loi de la brièveté dans la communication vocale des macaques n'est pas un artefact de l'analyse des durées moyennes des appels. *J. Quant. Linguist.* **20**, 209 - 217. (doi:10.1080/09296174.2013.799917)
34. Sole' RV. 2010 Genome size, self-organization and DNA's dark matter (Taille du génome, auto-organisation et matière noire de l'ADN). *Complexité* **16**, 20 - 23. (doi:10.1002/cplx.20326)
35. Bezerra BM, Souto AS, Radford AN, Jones G. 2011 Brevity is not always a virtue in primate communication. *Biol. Lett.* **7**, 23 - 25. (doi:10.1098/rsbl.2010.0455)
36. Cover T, Thomas J. 2006 *Elements of information theory*. New York, NY : Wiley.
37. Genty E, Clay Z, Hobaiter C, Zuberbu' hler K. 2014 Multi-modal use of a socially directed call in bonobos. *PLoS ONE* **9**, e84738. (doi:10.1371/journal.pone.0084738)
38. Sokal RR, Rohlf FJ. 1995 *Biométrie : les principes et la pratique des statistiques dans la recherche biologique*. New York, NY : WH Freeman & Co.
39. Ferrer-i-Cancho R, Herna'ndez-Ferna'ndez A, Baixeries J, De_bowski L, Ma'cutek J. 2014 Quand la loi de Menzerath-Altmann est-elle mathématiquement triviale ? Une nouvelle approche. *Stat. Appl. Genet. Mol. Biol.* **13**, 633 - 644. (doi:10.1515/sagmb-2013-0034)
40. Ferrer-i-Cancho R, Forns N, Herna'ndez-Ferna'ndez A, Bel-Enguix G, Baixeries J. 2013 The défis des modèles statistiques du langage : le cas de la loi de Menzerath dans les génomes. *Complexity* **18**, 11 - 17 (doi:10.1002/cplx.21429)
41. Herna'ndez-Ferna'ndez A, Baixeries J, Forns N, Ferrer-i-Cancho R. 2011 Size of the whole versus number de parties dans les génomes. *Entropy* **13**, 1465 - 1480. (doi:10.3390/e13081465)
42. Ridley DR. 1982 Zipf's law in transcribed speech. *Psychol. Res.* **44**, 97 - 103. (doi:10.1007/BF00308559)
43. Ferrer-i-Cancho R. 2016 Compression et les origines de la loi de Zipf pour les fréquences de mots. *Complexité* **21**, 409 - 411. (doi:10.1002/cplx.21820)
44. Held SDE, S'pinka M. 2011 Animal play and animal bien-être. *Anim. Behav.* **81**, 891 - 899. (doi:10.1016/j.anbehav.2011.01.007)
45. Ferrer-i-Cancho R. 2017 Le placement de la tête qui maximise la prévisibilité : une étude de l'impact sur la santé. Approche théorique de l'information. *Glottométrie* **39**, 38 - 71.
46. Ferrer-i-Cancho R. 2005 Zipf's law from a communicative phase transition. *Eur. Phys. J. B-Condens. Matter Complex Syst.* **47**, 449 - 457. (doi:10.1140/epjb/e2005-00340-y)
47. Pellis SM, Pellis VC. 1996 On knowing it's only play : the role of play signals in play fighting. *Aggress. Violent Behav.* **1**, 249 - 268. (doi:10.1016/1359-1789(95)00016-X)
48. Ko'hler R. 1987 Linguistique théorique systémique. *Theor. Linguist.* **14**, 241 - 258. (doi:10.1515/thli.1987.14.2-3.241)
49. Scott CB. 2008 *A primer for the exercise and nutrition sciences : thermodynamics, bioenergetics, metabolism*. Totowa, NJ : Humana Press.