

Origines du langage, évolutions linguistiques et hypothèses sur le langage Ummite : une lecture inspirée de Sverker Johansson

R Galli Mai 2026

La thèse de Sverker Johansson sur l'origine du langage et sa valeur scientifique

Référence : Johansson, S. (2021). *The dawn of language: The story of how we came to talk*. MacLehose Press.

Depuis des siècles, la question de l'origine du langage fascine philosophes, linguistes et biologistes. Comment une espèce animale a-t-elle pu devenir une espèce parlante ? Pourquoi les humains disposent-ils d'un système symbolique extraordinairement souple alors que les autres animaux, même très intelligents, ne possèdent pas d'équivalent direct ?

Dans son ouvrage consacré à l'origine du langage, Sverker Johansson propose une réponse qui se veut à la fois prudente et profondément ancrée dans les sciences de l'évolution. Son idée centrale est simple : le langage n'est pas apparu soudainement sous la forme d'un système complet. Il est le résultat d'une longue histoire où interagissent biologie, cognition, coopération sociale et évolution culturelle.

Johansson s'oppose ainsi aux modèles dits « saltationnistes », selon lesquels une mutation unique aurait produit presque instantanément une faculté grammaticale complète. Il défend au contraire une évolution progressive. Les premières formes de communication symbolique auraient été limitées, proches d'un « protolangage », avant de devenir les langues complexes que nous connaissons aujourd'hui.

L'un des aspects les plus originaux de sa démarche est de considérer les langues comme des systèmes évolutifs comparables, jusqu'à un certain point, aux espèces biologiques. Les langues changent, divergent, se transforment, disparaissent et se recombinent. Elles sont des objets culturels vivants.

Scientifiquement, cette approche est généralement appréciée pour son caractère interdisciplinaire. Johansson mobilise des données issues de la paléanthropologie, de la linguistique historique, des neurosciences, de la primatologie et de l'archéologie. Il ne prétend pas disposer d'une preuve définitive sur le « premier mot » prononcé par l'humanité ; il cherche plutôt à définir les contraintes auxquelles toute théorie sérieuse de l'origine du langage doit répondre.

Quelques exemples concrets : Néandertaliens, dispersion africaine et évolution des langues

L'un des arguments les plus intéressants de Johansson concerne les Néandertaliens.

Longtemps, on les a représentés comme des êtres cognitivement très inférieurs à l'humain moderne. Les découvertes récentes ont progressivement nuancé cette image. Les Néandertaliens utilisaient des outils élaborés, maîtrisaient le feu, pratiquaient certains comportements symboliques et possédaient un cerveau de grande taille.

Johansson considère donc plausible qu'ils aient disposé d'une forme de langage relativement développée. Il ne parle pas nécessairement d'un langage identique au nôtre, mais d'un système déjà avancé, bien antérieur à l'apparition de l'humain moderne.

Cette idée rejoint un autre aspect majeur : l'origine commune des populations humaines. Les modèles actuels de la paléanthropologie soutiennent qu'*Homo sapiens* provient d'Afrique avant de se disperser progressivement sur la planète. Si cette hypothèse est correcte, alors les milliers de

langues actuelles ne résultent pas de créations indépendantes surgies partout à la fois. Elles dériveraient d'une capacité linguistique commune qui s'est ensuite diversifiée au fil des migrations et des séparations sociales.

Le français constitue un exemple remarquable de cette évolution lente.

Vers l'an 1000, on rencontrait encore des formes comme *li reis* ou *li rois*. Peu à peu, les déclinaisons héritées du latin disparaissent. Les formes se simplifient, l'ordre des mots se stabilise et la langue devient plus proche du français moderne.

Le célèbre texte des Serments de Strasbourg permet d'observer cette transition. On y lit :

« Pro Deo amur et pro christian poblo et nostro commun salvament... »

Cette phrase reste reconnaissable, mais elle est déjà éloignée du français contemporain :

« Pour l'amour de Dieu et pour le peuple chrétien et notre salut commun... »

Entre ces deux états, il n'y a pourtant aucune rupture brutale. La langue change par petites modifications accumulées pendant des siècles. Pour Johansson, cette évolution graduelle constitue précisément le modèle général du changement linguistique.

Que penserait Johansson d'une civilisation extraterrestre avancée et de ses modes de communication ?

Les développements suivants concernant les extraterrestres ne figurent pas chez Johansson ; ils constituent une extrapolation personnelle inspirée de son cadre théorique.

Si l'on prolonge le raisonnement de Johansson, une question devient intéressante : une civilisation technologiquement avancée pourrait-elle exister sans langage ?

Johansson relie étroitement le langage à la coopération, à la mémoire collective et à la transmission culturelle. Une société capable de construire une technologie complexe, de transmettre des connaissances sur des générations et d'organiser une activité scientifique durable aurait probablement besoin d'un système symbolique puissant.

Cela ne signifie pas que ce système ressemblerait au langage humain.

Une espèce extraterrestre pourrait communiquer par signaux lumineux, impulsions électromagnétiques, substances chimiques ou interfaces biologiques inconnues. Rien n'impose des mots, des phrases ou même des sons.

Mais si cette civilisation possède une culture cumulative comparable à la nôtre, elle devrait vraisemblablement disposer d'un moyen stable de représenter et transmettre l'information.

Une analogie moderne aide à comprendre cette idée : Internet.

Notre civilisation ne repose plus sur un seul système symbolique. Elle empile des couches : langues naturelles, mathématiques, langages informatiques, protocoles de communication, notations scientifiques. Plus la technologie progresse, plus les systèmes symboliques se multiplient.

Une civilisation interstellaire pourrait fonctionner de manière analogue, avec des niveaux de communication très différents mais coordonnés.

Le langage Ummite de Jean Pollion dans le cadre de la théorie de Johansson

Le cas du langage Ummite tel qu'interprété par Jean Pollion est particulièrement intéressant lorsqu'on le confronte au modèle de Johansson.

Le système proposé par Pollion attribue une valeur conceptuelle élémentaire aux sons eux-mêmes. Les unités minimales ne seraient plus arbitraires : elles porteraient directement du sens.

Or cela entre en tension avec plusieurs idées centrales de Johansson.

D'abord, le modèle de Johansson est profondément évolutionnaire. Les langues réelles lui apparaissent comme le produit d'une longue histoire faite d'irrégularités, de ré analyses et de transformations successives. Elles ressemblent davantage à des organismes ayant accumulé des couches historiques qu'à des architectures conçues d'un seul bloc.

Le langage Ummite, au contraire, est souvent présenté comme remarquablement cohérent, presque parfaitement logique et étonnamment stable.

Cette perfection apparente soulève une difficulté : comment un tel système aurait-il émergé par évolution graduelle ?

Une autre question concerne la dynamique historique. Les langues humaines changent continuellement. Elles absorbent des emprunts, créent des ambiguïtés, développent des exceptions. Un système où chaque unité élémentaire posséderait un sens fixe pourrait devenir plus rigide face à l'innovation culturelle.

Dans une lecture personnelle mais inspirée de Johansson, cette stabilité pourrait suggérer non pas une langue naturelle, mais une langue construite, c'est-à-dire une *conlang*. Une langue artificielle peut rechercher la cohérence, éliminer les irrégularités et introduire une logique explicite. Ce type de structure est bien connu chez les humains.

Cela ne prouve évidemment rien sur le cas Ummite lui-même ; cela montre seulement qu'une langue fortement organisée n'implique pas automatiquement une origine non humaine.

Comparaison bayésienne de trois scénarios

Trois scénarios peuvent être comparés.

Le premier suppose que le décodage de Jean Pollion est essentiellement correct et qu'il décrit une langue extraterrestre authentique. Ce scénario exige plusieurs hypothèses simultanées : existence d'une civilisation extraterrestre, contact réel, authenticité documentaire et validité du décodage. Son coût explicatif est élevé.

Le second scénario imagine une langue artificielle extraterrestre élaborée par une réforme politique majeure : rupture générationnelle, reconstruction lexicale et conservation d'une écriture plus ancienne. Ce modèle résout certaines difficultés mais ajoute de nouvelles hypothèses. Il devient rapidement complexe et d'un coût supérieur au scénario précédent. Cette idée proposée par Alban Nanty est détaillée en Annexe 1.

Le troisième scénario considère le corpus comme une construction humaine : il s'agit d'une langue artificielle mais humaine cette fois-ci, c'est-à-dire un système élaboré volontairement ou conlang.

D'un point de vue bayésien, ce dernier scénario bénéficie d'un avantage très important : nous connaissons déjà des langues construites humaines. L'Espéranto, le Quenya ou le Lojban montrent qu'il est possible de créer des systèmes cohérents, parfois très sophistiqués.

Autrement dit, le scénario « conlang humaine » réutilise des phénomènes observés alors que les scénarios extraterrestres nécessitent de nombreuses hypothèses supplémentaires.

Cela ne signifie pas que les autres scénarios soient impossibles. Cela signifie seulement qu'ils paient un coût explicatif bien plus élevé comme je le détaille en Annexe 2.

Comment un bébé Ummite apprendrait-il une langue aussi structurée ?

Une question apparaît presque immédiatement lorsque l'on confronte le langage Ummite à la théorie de Johansson : même si l'on admet provisoirement que ce système existe, comment un enfant pourrait-il l'apprendre ?

Cette interrogation est importante parce que, chez Sverker Johansson, l'origine du langage ne se réduit pas à la structure des langues adultes. Une place essentielle est accordée au développement de l'enfant. Le langage humain est indissociable de l'apprentissage précoce, de l'attention conjointe, de l'imitation et de la coopération sociale.

Chez les humains, l'acquisition du langage suit généralement une progression allant du concret vers l'abstrait. Le bébé identifie des personnes, des objets, des actions simples. Les mots apparaissent progressivement. Les relations plus complexes viennent ensuite : temps, causalité, catégories générales, métaphores.

Le modèle Ummite proposé par Jean Pollion semble suivre une logique presque inverse. Les unités minimales ne seraient plus de simples sons arbitraires ; elles porteraient déjà une charge conceptuelle élémentaire. Les phonèmes eux-mêmes deviendraient signifiants.

Si l'on transpose cela au développement d'un enfant, une difficulté apparaît. Comment un nourrisson passerait-il de son expérience sensorielle immédiate à un système où les briques élémentaires seraient déjà abstraites ?

Deux possibilités peuvent alors être imaginées.

La première consisterait à supposer une cognition très différente de la cognition humaine. Les bébés Ummites disposeraient très tôt d'une capacité à manipuler des relations abstraites. Dans ce cas, le problème linguistique deviendrait en réalité un problème biologique : il faudrait expliquer une architecture cognitive nouvelle. Cette hypothèse est logiquement possible, mais elle nous éloigne du terrain empirique étudié par Johansson.

La seconde possibilité est plus compatible avec le cadre théorique de Johansson : le langage Ummite ne serait pas acquis directement sous sa forme « conceptuelle ». L'enfant commencerait par apprendre des mots concrets, des situations, des usages sociaux. Les règles profondes n'apparaîtraient que plus tard, un peu comme les humains découvrent progressivement la grammaire implicite de leur propre langue.

Dans cette perspective, le système conceptuel fonctionnerait davantage comme une couche savante, comparable à une analyse métalinguistique, plutôt que comme le mécanisme réel de l'apprentissage infantile.

Cette hypothèse présente un avantage important : elle conserve une trajectoire développementale proche de celle observée chez les humains. Le bébé apprendrait d'abord le monde ; la structure profonde viendrait ensuite.

Une analogie moderne peut aider à comprendre cette idée. Un enfant humain parle couramment bien avant de connaître les règles de grammaire. Il sait utiliser le français sans avoir étudié les notions de sujet, complément ou subordonnée. De la même manière, un enfant Ummite pourrait utiliser sa langue avant de comprendre les règles conceptuelles censées la sous-tendre.

Cette réflexion conduit à une conséquence intéressante. Si le langage Ummite exige effectivement une connaissance explicite de ses relations internes dès les premières étapes du développement, alors il paraît difficile à concilier avec le modèle de Johansson. Si au contraire ces relations n'apparaissent qu'après coup, le système se rapproche davantage d'une langue naturelle enrichie d'une couche analytique.

On retrouve ici une tension déjà rencontrée plus haut : le langage Ummite semble souvent décrit comme un système très ordonné, presque achevé, alors que Johansson voit dans les langues réelles des objets historiques, irréguliers et façonnés par l'apprentissage progressif.

La question du bébé devient alors un test inattendu. Une langue n'est pas seulement ce qu'un adulte savant peut analyser ; elle est aussi ce qu'un enfant peut acquérir spontanément.

Et c'est peut-être sur ce terrain, plus encore que sur la logique interne du système, que Johansson aurait cherché selon moi des réponses. Je propose ma vision en Annexe3.

Conclusion

La théorie de Sverker Johansson propose une vision du langage profondément ancrée dans l'histoire naturelle. Le langage humain n'est ni une invention soudaine ni une structure tombée du ciel : il résulte d'une longue interaction entre évolution biologique, coopération sociale et transformation culturelle.

Cette perspective éclaire non seulement les langues humaines connues, mais aussi des questions plus spéculatives. Elle permet d'interroger la possibilité d'un langage extraterrestre, la nature de systèmes symboliques très organisés ou encore le statut de constructions comme le langage Ummite.

Face à ces hypothèses, Johansson conduirait probablement à poser une question simple mais exigeante : *quelle est l'histoire évolutive du système étudié ?*

Car pour lui, une langue n'est pas seulement une structure logique. C'est une histoire accumulée, transmise et transformée par des générations de locuteurs.

Une langue est aussi une trajectoire d'apprentissage. Une théorie du langage doit expliquer non seulement comment les adultes communiquent, mais aussi comment un enfant entre progressivement dans ce système. Dans une lecture inspirée de Johansson, le problème du développement devient alors un critère essentiel pour évaluer des systèmes très structurés comme le langage Ummite.

Annexe 1 Le scénario d'Alban Nanty tel qu'il me l'a proposé dans un mail le 18/01/2026

1) **Je pense que la langue oummaïne est une langue artificielle** (au sens que les linguistes donnent à ce terme, par opposition à "langue naturelle"). Des langues artificielles, les linguistes terriens en ont créées plein, la plus connue étant l'espéranto qui possède quelques milliers de locuteurs. D'autres ont été créées pour des œuvres de fictions par exemple le Klingon, créé pour la série Star Trek, parlé par quelques centaines de fans de la série, et qui la pratiquent durant des conventions de fan. On peut aussi citer l'elfique créé par Tolkien pour son univers du seigneur des anneaux. Toutes ces langues sont dites artificielles car construites, par opposition aux langues naturelles qui émergent d'un groupe d'individu. A la préhistoire, un groupe d'individu désigne un objet et font un "grognement" pour assigner un "mot" à cet objet. Puis petit à petit la langue se structure, mais de façon organique. Donc les langues naturelles ne possèdent pas de structure logique/mathématiques, et sont généralement bourrées d'exception grammaticale.

Il est probable que le système de soncept découvert par Pollion soit une réalité, mais si c'est le cas, alors pour moi cela signifie que la langue oummaïne est artificielle et non naturelle.

2) **Mise en application.** Si l'espéranto est bien meilleur que les langues naturelles (car par exemple elle ne souffre d'aucune exception, tous les noms communs finissent en "o", etc....) pourquoi n'est-elle pas parlée par tout le monde? Car il faut une **volonté politique**. Par exemple si la commission européenne votait une loi obligeant tous les pays européens à enseigner l'espéranto dans toutes les écoles de l'UE, alors peut-être que dans quelques génération la majorité des européens parlerait l'espéranto, et les langues nationales dépériraient jusqu'à disparaître. Sans volonté politique, aucune chance qu'une langue artificielle soit adoptée par un groupe de personnes, même si de toute évidence la langue artificielle est bien meilleure (puisque construite, par opposition à l'anarchie d'évolution d'une langue naturelle). Il me semble que sur Ummo, la volonté politique a été présente à l'époque de la "génération sacrifiée", c'est à dire lorsque qu'Ummoalewee a décidé de séparer tous les enfants d'Ummo de leurs parents pour les élever dans les écoles sans contact possible avec les parents. Ce cadre permet de ne parler et d'instruire ces enfants que dans la langue artificielle (les professeurs devant faire l'effort de l'apprendre). En une génération seulement la langue artificielle peut s'imposer dans la population.

3) **Le langage Ummite possède des mots-objets.** Sur ce point je diffère complètement de l'idée de Pollion (que l'on trouvera dans son livre). Je pense que les linguistes ont sans doute édicté des règles sonceptuelles pour créer les mots de la nouvelle langue, mais qu'ensuite ils ont appliqué leurs règles pour recréer tout le dictionnaire. Par exemple là où un fruit juteux s'appelait "tartenpion" dans la langue naturelle, ils l'ont appelé "inowi" dans la nouvelle langue. Mais ce nouveau mot créé à la sauce sonceptuelle désigne bien le même objet, et désignera toujours le même objet. Bien sûr quelques variations du mot sont possibles, tout comme nous le faisons aussi avec nos langues flexionnelles tel que l'allemand. Ainsi on pourra rajouter un "i" à la fin pour indiquer le pluriel (j'expliquerai le pluriel un autre jour). Je n'adhère pas du tout à l'idée de Pollion qui croient que les Ummites n'ont pas de dictionnaire (je pourrais élaborer plus tard sur ce point), même si je reconnais que la langue Ummite permet très facilement la création de néologisme, c'est normal, elle a été conçue pour cela (et au passage, c'est pourquoi les Ummites créent parfois des néologismes avec notre langue, car ils ont l'habitude de le faire tout le temps dans la leur).

4) **L'écrit.** A l'époque de la création de la langue, les Ummites avaient déjà perdu l'habitude d'archiver des textes, ils n'archivent que de la vidéo ou de l'audio. Ainsi les linguistes n'ont pas jugé nécessaire d'inventer des signes pour chaque soncept... ou bien même s'ils l'ont fait (car l'effort à fournir pour le faire est négligeable, donc c'est probable qu'ils l'aient fait quand même), personne n'a

besoin de l'utiliser. Mais ils possédaient beaucoup d'archives écrites dans l'ancien système d'écriture qui ressemble à ce qu'on fait dans l'écriture chinoise. Ainsi, ils ont gardé cette écriture, simplement là où un ancien prononçait un idéogramme avec le mot de la langue naturelle, un contemporain prononçait le même idéogramme avec la prononciation sonceptuelle. C'est pourquoi, je pense que l'écriture du langage Ummite reste identique à ce qu'elle était lorsqu'ils parlaient encore leur langage naturel, et qu'elle est composée d'un ensemble d'idéogrammes, n'ayant aucun rapport avec le système sonceptuel.

Annexe 2 Comparaison bayésienne des 3 scenarii

Le rasoir d'Ockham comme principe méthodologique

Le raisonnement scientifique ne consiste pas à imaginer tous les scénarios possibles, mais à déterminer lesquels sont rationnellement préférables à partir des données disponibles. C'est dans ce cadre que s'inscrit le principe de parcimonie, traditionnellement associé à Guillaume d'Ockham et résumé par la formule latine *Entia non sunt multiplicanda praeter necessitatem*, que l'on traduit généralement par : « Il ne faut pas multiplier les entités (ou les nouveaux concepts) sans nécessité ».

Ce principe ne stipule pas que l'hypothèse la plus simple est nécessairement vraie, mais qu'à pouvoir explicatif équivalent, une hypothèse plus complexe est pénalisée tant qu'elle n'apporte pas un gain explicatif décisif. En termes modernes, et notamment dans le cadre de l'analyse bayésienne, la parcimonie n'est pas une préférence esthétique pour la simplicité, mais une conséquence mathématique du fait que les hypothèses complexes répartissent leur probabilité a priori sur un plus grand nombre de paramètres et de configurations possibles.

Appliqué au cas du langage Ummite, ce principe invite à comparer non pas ce qui est imaginable, mais ce qui est le plus probable compte tenu des données linguistiques disponibles.

Trois ensembles d'hypothèses

Le premier ensemble suppose que le décodage de Pollion est essentiellement correct et qu'il décrit une langue extraterrestre authentique. Ce scénario dit JP exige plusieurs hypothèses simultanées :

1. Existence d'une civilisation extraterrestre.
2. Capacité technologique à voyager dans les étoiles
3. Contact réel sur Terre.
4. Authenticité documentaire des lettres Ummites
5. Validité du décodage de Jean Pollion.

Le second scénario dit AN imagine de plus une langue artificielle extraterrestre élaborée par une réforme politique majeure sur Ummo :

1. Le langage Ummite serait une langue artificielle réelle, conçue de manière délibérée, par opposition à une langue naturelle issue d'une évolution organique. Cette langue artificielle aurait été imposée à l'ensemble de la population par une volonté politique forte, à la faveur d'une rupture générationnelle (séparation des enfants et des parents).
2. Le langage posséderait un dictionnaire stable de mots-objets, les règles sonceptuelles ayant servi à recréer systématiquement le lexique à partir d'une langue naturelle antérieure.
3. Une écriture idéographique ancienne aurait été conservée, indépendante du système phonologique sonceptuel, expliquant l'absence de correspondance observable entre écriture et lexique transmis.

Ce modèle résout certaines difficultés mais **ajoute trois nouvelles hypothèses** à celles de Jean Pollion car on conserve ici l'idée d'une source ET.

Le troisième scénario dit RG, que j'analyserai dans un article indépendant, considère le corpus comme une construction humaine, hypothèse bien plus modeste :

1. Le corpus Ummite constitue un langage artificiel ou un système lexical construit, mais il ne présente pas les propriétés observables d'une langue fonctionnelle complète (grammaire, flexion, contraintes syntaxiques).

2. Il peut être expliqué de manière satisfaisante par une production humaine centralisée, sans qu'il soit nécessaire de postuler une langue réellement pratiquée par une civilisation entière.

Dans une perspective bayésienne, il ne s'agit pas de déterminer si tel ou tel scénario est possible, mais d'évaluer son coût explicatif par rapport à des hypothèses concurrentes.

Prenons par exemple le scénario AN. Ce scénario suppose simultanément : l'existence d'une civilisation ET, sa présence sur Terre, une communication épistolaire. A cela, il faut ajouter la conception d'une langue artificielle pleinement fonctionnelle ; son imposition réussie à une population entière ; une transition linguistique sans traces observables ; l'absence de grammaire détectable dans le corpus transmis ; la conservation d'un système d'écriture indépendant, non documenté ; une transmission fragmentaire parfaitement sélective vers la Terre.

Chacune de ces hypothèses prise isolément est envisageable. **Mais leur conjonction multiplie les entités explicatives sans produire de prédictions linguistiques supplémentaires vérifiables.** En termes bayésiens, chaque hypothèse supplémentaire dilue la probabilité a priori du modèle global.

À l'inverse, l'hypothèse d'une production humaine centralisée du corpus ne requiert qu'un nombre restreint d'hypothèses : une intention, une cohérence cognitive, un cadre culturel donné. Elle explique tout aussi bien les données linguistiques observées, sans nécessiter de mécanismes additionnels non testables.

Le scénario dit JP se situe entre les ces deux scenarii en termes de probabilités : conformément au principe de parcimonie, et à pouvoir explicatif équivalent, l'hypothèse RG est de loin la plus parcimonieuse.

Traduction probabiliste : $P[RG] \gg P[JP] \gg P[AN]$

Annexe 3 Un bébé Ummite peut-il apprendre l'Ummite ?

Cette annexe s'appuie sur le livre suivant que je conseille vivement au lecteur :

Lê Nguyễn Hoang et al (2018). *La formule du savoir: une philosophie unifiée du savoir fondée sur le théorème de Bayes*. EDP Sciences.

Je me suis demandé en particulier à la lecture du chapitre 19 de ce livre si un bébé Ummite, dont j'ai supposé le cerveau identique à un bébé humain, serait capable de repérer des régularités statistiques dans le langage Ummite, des segmentations, combiner des mots pour faire des phrases, etc., le tout en s'appuyant sur le lexique de J Pollion.

Dans ce qui suit, le terme « langage Ummite » désigne exclusivement le corpus lexical attesté, et non une langue naturelle ou fonctionnelle au sens linguistique strict.

Synthèse du chapitre 19 sur l'apprentissage du langage

Le chapitre 19 établit plusieurs points robustes concernant l'apprentissage du langage chez le bébé humain :

1. Le bébé dispose très tôt de capacités statistiques implicites.
2. Il est capable de
 - Détecter des régularités statistiques dans les flux sonores ;
 - Segmenter des suites continues en unités lexicales probables ;
 - Inférer des catégories grammaticales (nom, verbe, adjectif) à partir de la distribution des mots ;
 - Généraliser le sens de mots nouveaux à partir de très peu d'exemples ;

Ces capacités sont modélisables par une inférence bayésienne hiérarchique, reposant sur des a priori structuraux (prédispositions cognitives), et une mise à jour probabiliste par exposition répétée aux données.

Un point crucial est explicitement formulé : le bébé apprend à partir de données riches, redondantes, distribuées dans le temps, et principalement à partir de phrases complètes, pas de listes de mots isolés.

Ce qu'un bébé bayésien attend d'un langage

Pour que l'apprentissage bayésien du langage fonctionne, certaines conditions minimales sont requises. Il faut chez les humains :

- Un flux continu de parole ;
- Des cooccurrences stables entre unités ;
- Une syntaxe observable, même rudimentaire ;
- Des variations contextuelles permettant d'inférer rôles et relations ;
- Une pression communicative réelle (interaction sociale).

Ces éléments ne sont pas accessoires : ils constituent les données sur lesquelles l'inférence bayésienne opère.

Application au lexique Ummite de Jean Pollion

On suppose qu'un bébé Ummite possède à minima les mêmes capacités cognitives qu'un bébé humain.

Sous cette hypothèse, on peut se poser la question suivante :

Le corpus Ummite tel que nous le connaissons fournirait-il des données suffisantes à un cerveau bayésien pour apprendre une langue ?

La réponse est négative, pour les raisons suivantes :

1. Absence de flux linguistique : nous ne disposons pas de corpus de parole continue, mais d'un lexique isolé.
2. Absence de phrases complètes exploitables. Or le chapitre 19 insiste sur le fait que la segmentation lexicale et grammaticale repose sur la structure des phrases.
3. Absence de grammaire observable : en l'absence de données permettant d'inférer une grammaire, celle-ci est linguistiquement non testable. Sans ordre syntaxique stable, sans flexion, sans particules, un cerveau bayésien ne peut pas inférer de catégories grammaticales.
4. Absence de distribution statistique exploitable. Les « concepts » sont définis a posteriori par l'analyste, non appris par exposition répétée.

Un bébé bayésien, humain ou Ummite, ne pourrait pas apprendre une langue fonctionnelle à partir du matériau Ummite tel qu'il est attesté.

Ce que cela implique pour la théorie de Jean Pollion

Le système de concepts de Jean Pollion est très insuffisant pour évaluer sa compatibilité avec un modèle bayésien d'acquisition du langage tel que décrit dans le chapitre 19. Mais on peut remarquer que :

- Les concepts ne sont pas découverts par apprentissage statistique,
- Ils sont imposés comme primitives explicatives,
- Ils ne génèrent ni syntaxe, ni prédictions testables indépendantes.

Autrement dit, le système de Pollion fonctionne comme une clé de décodage interprétative, pas comme un langage issu de modèles bayésiens de l'apprentissage.

Conclusion

À la lumière des travaux contemporains sur le « cerveau bayésien » et l'acquisition du langage chez le bébé, le lexique Ummite tel qu'il est présenté par Jean Pollion ne fournit pas les données minimales permettant l'apprentissage d'une langue par inférence statistique. En l'absence de flux linguistique, de syntaxe observable et de régularités distributionnelles exploitables, un cerveau dit bayésien (humain ou supposément Ummite) ne pourrait ni segmenter, ni généraliser, ni produire des phrases.

Le système des concepts apparaît dès lors non comme le produit d'un apprentissage linguistique naturel, mais comme une construction interprétative a posteriori.