

Based on a talk given at "L'école d'été de la linguistique française", Agay (Côte d'Azur), September 2012.

Esa Itkonen (Université de Turku, La Finlande) & Anneli Pajunen (Université de Tampere, La Finlande)

## LA LINGUISTIQUE EST-ELLE UNE SCIENCE DÉDUCTIVE? – POUR LA PLÛPART, NON!

### 1) Note préliminaire

La linguistique n'est pas une discipline monolithique. C'est plutôt un ensemble qui comporte un bon nombre de méthodes différents. À première vue, leur diversité peut être déconcertante. Mais si l'on y regarde de plus près, les méthodes s'arrangent dans un ordre bien défini, c'est-à-dire un ordre qui se qualifie de "présuppositionnel". Au lieu d'aborder ce sujet en termes généraux, nous préférons donner tout de suite un exemple qui fait comprendre d'une façon concrète de quoi il s'agit.

### 2) Expérimentation

Les études de la fixation des yeux (*eye-fixation studies*) représentent l'expérimentation dans le sens strict du mot en ce que les processus étudiés se situent tout à fait en dehors du contrôle volontaire. La simplicité ou bien la complexité des constructions grammaticales est mesurée en découvrant combien de temps il faut pour les lire et, plus spécialement, quels sont les mots qui forcent le regard à revenir en arrière. C'est de cette façon qu'on peut étudier, par exemple, la nature de l'**accord grammatical** entre le nom et l'adjectif déterminant. Les langues diffèrent, bien sûr, à cet égard. En Latin, par exemple, il y a un accord en genre, nombre et cas. En français, il y a, au moins dans la langue écrite, un accord en genre et nombre. En finnois, il y a un accord en nombre et cas. En anglais, il n'y a d'accord ni en genre ni en nombre ni en cas.

Ce qui suit se base sur Vainio, Hyönä & Pajunen (2003, 2008, 2009). Ici, l'enquête se porte sur le finnois, une langue où l'accord s'exprime par des suffixes et où – ce qui est important – l'adjectif précède le nom qu'il détermine. Dans ce contexte il faut signaler trois constructions différentes: (1) le nom seul; (2) l'adjectif qui s'accorde et le nom; (3) l'adjectif indéclinable (ou sans accord) et le nom:

(1)        \_\_\_ N-x

- (2) A-x N-x
- (3) A N-x

- (1) ... \_\_\_\_\_ orkesteri-ksi ...
- (2) ... mainio-ksiorkesteri-ksi ...
- (3) ... kelpo-∅ orkesteri-ksi ...

- (1) ... \_\_\_\_\_ orchestre-cas&nombre ...
- (2) ... bon-cas&nombre orchestre-cas&nombre
- (3) ... brave-∅ orchestre-cas&nombre

On a formulé des hypothèses très précises concernant les facteurs qui déterminent la durée de la lecture dans chaque cas. Mais ici, pour simplifier, nous donnons seulement le résultat de cette expérience, sur une échelle qui va de la plus courte durée à la plus longue:

$$(2) < (1) < (3)$$

Voici un résultat assez surprenant: on lit plus vite deux mots qui s'accordent qu'on lit un seul mot. C'est-à-dire, un seul regard suffit pour qu'on comprenne la construction (2), tandis que, pour comprendre la construction (1) dans le contexte plus large de la phrase, il faut que le regard revienne sur le mot unique. Ce qui est moins surprenant, c'est qu'on lit le plus lentement la construction (3), c'est-à-dire deux mots qui, à l'encontre de la règle générale, ne s'accordent pas.

La signification de cette expérience, c'est qu'il justifie l'existence de l'accord grammatical, en démontrant qu'il facilite la compréhension. Ce résultat réfute l'opinion générale qui a été formulée par John Haiman (1985: 164) de la façon suivante: "Grammatical agreement is redundant: not only non-iconic but meaningless" (= "L'accord grammatical est redondant: non seulement non-iconique mais aussi dénué de sens"). Voilà une opinion qui est – a priori – tout à fait plausible. Mais la raison d'être même de l'enquête expérimental, c'est prouver que, de temps en temps, nos opinions aprioristes, aussi plausibles soient-elles, manquent de bien-fondé. Et c'est justement ce que l'expérience ici décrite a accompli, en donnant une **explication fonctionnelle** à l'accord grammatical. Pourquoi explication fonctionnelle? Parce que l'accord s'avère être un moyen pour atteindre le but de la compréhension aisée.

### 3) Une hiérarchie des méthodes empiriques

On vient de voir un exemple de l'expérimentation au sens strict du mot. Mais ce n'est que maintenant que, à proprement parler, notre histoire commence. Il faut poser ces questions cruciales: Qu'est-ce qui s'est passé **avant** que l'expérience proprement dite soit faite? Qu'est-ce que l'expérience **présuppose**-t-elle?

Ici, les mots cibles sont l'adjectif déclinable *mainio* ('bon') et l'adjectif indéclinable *kelpo* ('brave'), et on vient de voir qu'ils sont lus différemment. Mais cette différence de lecture ne peut être attribuée à leur différence grammaticale qu'à condition que toutes les autres explications ont été exclues au préalable. C'est-à-dire, les deux mots doivent être comparables sous tous les autres aspects. Cela est assuré par deux méthodes différentes.

Tout d'abord, les deux mots doivent être tout aussi fréquents, ce qui exige qu'on compte combien de fois ils se rencontrent dans un **corpus** suffisamment étendu.

Ensuite, il faut vérifier que les deux mots possèdent exactement les mêmes valeurs stylistiques, ce qui exige qu'on fasse remplir à un groupe de sujets des **questionnaires** ci-concernant; et cela est vrai non seulement des deux mots mais aussi des phrases d'épreuve (*test sentences*) qui les contiennent.

C'est seulement à la suite de ces stades préliminaires qu'on peut se mettre à faire l'expérience proprement dite. À ce moment-ci, notre argument se résume dans la hiérarchie suivante:



Sur cette échelle, qui va de B à D, l'ordre temporel coïncide avec l'ordre logique ou présuppositionnel: le stade B précède les autres et peut exister sans eux; le stade C présuppose le stade B mais non pas le stade D. Le stade D présuppose les autres. Cette échelle représente – en gros – le domaine de la **linguistique empirique**, c'est-à-dire l'étude des phénomènes langagiers qui se situent dans l'espace-temps et qui sont, par conséquent, accessible à l'observation, ou bien à l'ouïe ou bien à la vue. Les mêmes méthodes sont analysées par Gonzales-Marquez et al. (2007), mais sans être mis dans un ordre quelconque.

#### 4) Empirique vs. non-empirique

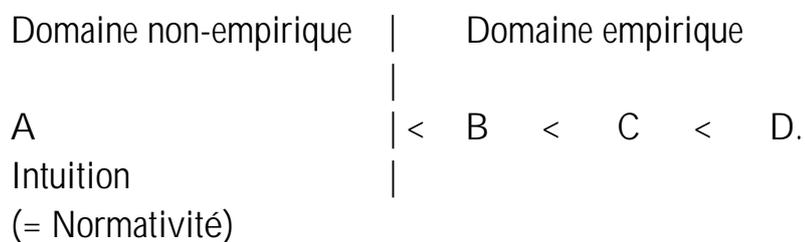
Le domaine de la linguistique empirique, est-ce que c'est la même chose que le domaine de la linguistique tout court? On est tenté de répondre 'Oui'. Nous entendons ce qui est dit ou nous voyons ce qui est écrit, et puis nous le décrivons,

voilà tout. Quoi de plus naturel? Et pourtant, il suffit d'un tout petit effort intellectuel pour convaincre n'importe qui de ce que cela ne peut pas être juste. Pourquoi pas? Parce que c'est beaucoup trop simple.

Les mots cibles *mainio* et *kelpo* jouent ici un rôle capital. Avant de compter combien de fois ces mots se rencontrent dans un corpus quelconque, il faut savoir ce qu'elles signifient et, surtout, il faut savoir que ce sont des mots de la langue finnoise, et non pas, par exemple, de la langue française ou anglaise. On possède ce type de savoir avant de s'attaquer à n'importe quel corpus; et c'est par le moyen de l'**observation** que se fait l'étude d'un corpus. Donc, le savoir en question qualifie, littéralement, de **pré-observationnel** (ou de **pré-expérimental**). Il est de coutume de le désigner sous le nom d'**intuition**, plus précisément, intuition linguistique. Il faut retenir cette différence très nette entre intuition et observation.

Ce n'est pas seulement les mots cibles qui nécessitent un emploi de l'intuition, mais également les phrases d'épreuve dont ces mots font partie. Ces phrases n'ont été observées dans aucun corpus, mais ils ont été construites par le linguiste, ce qui veut dire qu'ils se fondent sur son intuition à lui ou à elle. Il va sans dire que les phrases d'épreuve doivent être grammaticalement correctes, c'est-à-dire conformes aux **normes** de la langue, ici de la langue finnoise. C'est donc l'intuition qui donne accès aux normes langagières.

En somme, voici l'hierarchie des méthodes linguistiques dans son intégralité:



L'argument plus détaillé concernant la nature du stade A se trouve dans Itkonen (1978, 2003).

##### 5) Les filtres normatifs dans la linguistique de corpus

Maintenant il nous faut préciser de plus près la relation entre A et B, c'est-à-dire intuition et corpus; et, pour ce faire, nous choisissons un autre exemple. L'enquête qui va être rapportée ici se porte sur le changement dans le lexique des verbes en finnois pendant le dernier demi-siècle (voir Pajunen 2006). Pour établir la situation d'aujourd'hui, on se base sur les annuaires 2000 et 2001 de *Helsingin Sanomat* (= le journal quotidien le plus répandu de la Finlande), qui constituent un corpus de 31

millions de mots. Dans ce corpus, le nombre des verbes, c'est-à-dire des lexèmes verbaux, est exactement 8'011.

La question suivante se pose maintenant: comment, à partir du nombre 31'000'000, a-t-on arrivé au nombre 8'011? Au premier abord, cela paraît évident: on n'a que compter les verbes, et il se trouve qu'il y en a 8'011. Mais, comme on verra, ce n'est pas évident du tout. Pour éclairer le caractère de notre exemple nous nous servirons de la notion de **filtre normatif**. En effet, il y a toute une série de filtres normatifs qui doivent être traversés par le corpus de 31 millions d'occurrences de mots, pour qu'on arrive finalement à 8'011 lexèmes verbaux (voir Itkonen & Pajunen 2010: 29–32).

Les données primaires, ce sont ici les textes d'un journal quotidien. Quand on écrit un article pour qu'il soit publié dans un journal, il est normal qu'on se corrige de temps en temps. Ces corrections reposent sur l'intuition linguistique, c'est-à-dire sur le savoir qu'on a des normes de sa langue maternelle. Voici le premier filtre normatif.

Ensuite, le texte est imprimé. Jadis, il y avait un métier de "correcteur ou correctrice d'épreuves" qui consistait à vérifier les textes qui allaient être imprimés. Aujourd'hui cette besogne est exécutée par des programmes informatiques. Dans notre cas, il se trouve que le programme utilisé par le journal s'appelle FINTWOL. Il va sans dire que ce programme repose sur l'intuition linguistiques de ceux qui l'ont conçu. Voici le second filtre normatif.

À partir de ce moment-là, le corpus est à la disposition du linguiste. Au stade suivant, le corpus tout entier est analysé par un programme qui a été conçu pour des raisons purement scientifiques. Une description morpho-syntaxique est attribuée à chaque mot et, plus spécialement, une forme de base est attachée à toutes les occurrences d'un seul et même verbe. Dans notre cas, le programme analytique a été fait sur le modèle de la "Grammaire de Dépendance Fonctionnelle" (*Functional Dependency Grammar*) et repose, évidemment, sur l'intuition linguistique de ses créateurs. Voici le troisième filtre normatif.

Le programme analytique ne fonctionne pas toujours d'une manière parfaite. En particulier, lorsqu'il s'agit des verbes tout à fait nouveaux et qui, pour ainsi dire, hésitent à entrer dans la langue, le programme est susceptible de donner des formes de base impropres. Le travail de correction est exécuté d'abord à l'aide d'un programme informatique additionnel et ensuite d'une façon purement manuelle, ce qui s'avère très laborieux. En corrigeant les fautes, le linguiste se réfère à sa propre intuition linguistique. Voici le quatrième filtre normatif.

Finalement, il faut réviser les données une fois de plus. Maintenant il ne s'agit plus de corriger des fautes proprement dites mais d'éliminer ce qui est superflu. Les verbes dérivés dont la manière de dérivation est tout à fait productive et transparente

n'ont pas besoin d'être notés en plus du verbe simple qui est à leur origine. De même, ce n'est pas la peine de noter séparément toutes les variantes dialectales d'un seul et même verbe. Voici le cinquième filtre normatif.

En somme, il faut que, en cours d'être analysé, le corpus traverse au moins **cinque** filtres normatifs; et il serait aisé d'en distinguer encore plus.

#### 6) Les conséquences méthodologiques

Il se peut qu'on ne comprenne pas tout de suite l'importance méthodologique que les filtres normatifs possèdent pour la linguistique et, plus généralement, pour les sciences de l'homme. On dira que dans les sciences naturelles aussi, on est constamment obligé à faire des corrections. Cela est vrai, bien sûr, mais il y a pourtant une différence capitale vis-à-vis de ce qui vient d'être présenté ici. Dans les sciences naturelles, les **fautes** ou **erreurs** sont toujours commises par ceux qui étudient, jamais par ce qui est étudié. Dans les sciences humaines, par contre, elles sont commises – et corrigées – tout aussi bien par ceux qui sont étudiés que par ceux qui étudient. Bien sûr, il y a des **fautes de linguistique**, c'est-à-dire des fautes de nature théorique, qui sont inconnus au sujet parlant moyen. Mais l'important, c'est que le linguiste professionnel et le sujet parlant moyen partagent le même savoir qui les autorisent tous les deux à reconnaître les **fautes de langue** comme telles. C'est dans ce sens précis que l'intuition linguistique n'a aucun équivalent dans les sciences naturelles. Ceci explique aussi l'importance qui s'attache à la notion de "filtre normatif".

Le corollaire en est le suivant. Dans les sciences naturelles les normes, étant celles du scientifique, sont le moyen pour étudier ce qui se passe dans l'espace-temps; elles ne sont pas elles-mêmes l'objet d'étude. (Notons l'existence de la 'protophysique' développée par Paul Lorenzen; mais la protophysique n'est pas une science naturelle.) Dans la linguistique, par contre, les normes sont ou bien celles du scientifique ou bien celles partagées par le scientifique et le sujet parlant moyen; et ce sont les normes de ce dernier type, c'est-à-dire le stade A de notre hiérarchie, qui constituent l'objet d'étude primaire. Bien sûr, on étudie aussi ce qui se passe dans l'espace-temps, et ce sont les stades B, C et D, mais tout cela est de nature secondaire. Voilà la différence. (Il faut signaler que 'primaire' et 'secondaire' sont ici des termes purement techniques, sans aucun jugement de valeur).

#### 7) La nature de la linguistique intuitionnelle ou non-empirique

Il reste à répondre aux questions suivantes: Quelle est la nature de la méthode qui est employée dans le domaine de la linguistique non-empirique ou intuitionnelle, et comment sont les résultats produits par cette méthode-là? À notre avis, c'est la notion de "grammaire optimale", envisagée par Katz (1981), qui fournit ici le meilleur point de départ. Le but de cette sorte de grammaire, tout d'abord, c'est se concentrer sur toutes les phrases, et rien que celles-là, d'une certaine langue et les décrire d'une manière qui est à la fois aussi complète que possible et indépendante des limitations physiques, psychologiques ou sociologiques auxquelles les sujets parlants puissent être soumis. Il est possible que ces critères soient remplis par plusieurs grammaires en même temps. Donc, pour choisir la grammaire optimale, il faut un critère de plus, et c'est la **simplicité**: on choisit la grammaire qui est – ceteribus paribus – la plus simple.

Or, l'idéal **axiomatique**, c'est déduire le maximum de données d'une manière aussi simple que possible, c'est-à-dire, en utilisant le minimum d'axiomes, de définitions et de règles d'inférence. La grammaire optimale de Katz (1981) s'avère être un plaidoyer en faveur de la linguistique axiomatique ou **déductive**. Voilà la justification du terme "linguistique déductive" qui figure dans le titre de notre exposé. Étant indépendante des faits physiques, psychologiques et sociologiques, la linguistique déductive est dépourvue de toute notion de causalité.

Ce qui vient d'être dit fait au moins entrevoir la justification de l'analogie entre la linguistique intuitionnelle et la logique formelle. Si l'on veut connaître la nature d'une entité A, il ne faut pas la comparer à une seule entité B. Si l'on cherche des similarités entre A et B, on ne manquera pas de les trouver. C'est ainsi que la linguistique (conçue comme une discipline monolithique) n'a été comparée qu'à la physique (ou la biologie), et on a forcément conclu que les deux sont "similaires". Mais ce procédé ne vaut rien. Si l'on veut connaître la nature de la linguistique intuitionnelle, ce qu'il faudrait faire, c'est la comparer à la physique, d'une part, et (par ex.) à la logique déontique, de l'autre. Mais cela n'a jamais été fait, à l'exception de Itkonen (1975, 1978).

À notre avis, c'est la grammaire de Panini, composée il y a à peu près 2'400 ans, qui dans l'histoire de la linguistique se rapproche le plus de la 'grammaire optimale' (voir Itkonen 1991). Parmi ceux qui ont suivi Panini en préconisant la simplicité comme le critère fondamentale, il faut noter au moins Louis Hjelmslev, Zellig Harris et Noam Chomsky.

## 8) La nature de la linguistique empirique

Il convient d'approfondir un peu la nature de la linguistique empirique. Bien sûr, il s'agit d'étudier le comportement langagier dans l'espace-temps, mais cela ne suffit pas comme caractérisation. À notre avis, il y a quelque chose qui unit toutes les variantes de la linguistique empirique, et c'est l'**explication fonctionnelle**: l'homme agit ou s'abstient d'agir parce qu'il croit – à tort ou à raison – que ce qu'il fait ou s'abstient de faire sert de moyen pour atteindre un but quelconque. Ceci est la seule concept de **causalité** propre à la linguistique (voir Itkonen 1983).

Plus on s'enfonce dans les détails de n'importe quel type de linguistique empirique, plus on s'aperçoit que c'est toujours la même chose: l'homme cherche à obtenir ce qu'il désire en employant les moyens qu'il prend pour être dictés par la situation où il se trouve (voir Itkonen & Pajunen 2010). C'est une formule assez simple et facile à comprendre. Ce qui est difficile, c'est montrer comment elle se manifeste à travers le panorama énorme des faits de langage.

## 9) Conclusion

Les deux types d'enquête linguistique, non-empirique et empirique, visent aux buts propres à eux, c'est-à-dire présentation axiomatique, d'une part, et explication causale (voire fonctionnelle), de l'autre. L'histoire de la linguistique occidentale se caractérise par une tension confuse entre ces deux buts pourtant bien distincts (voir Itkonen 1991, 2013).

La typologie des langues est une entreprise **inductive** par excellence (voir Itkonen 2011), ce qui nous force à qualifier ce qui vient d'être dit concernant la nature du stade A de notre hiérarchie. Le point de vue dit déductif ne vaut que pour une langue qu'on connaît à fond (typiquement, mais pas seulement, la langue maternelle). Cela est vrai, bien sûr, des grands maîtres de l'histoire universelle de la linguistique: Panini, Tolkaappiyanaar, Apollonios Dyscole, Sibawaihi (voir Itkonen 1991, 2000).

## Bibliographie

- Gonzales-Marquez, Monica et al. (éds). 2007. *Methods in cognitive linguistics*. Amsterdam: Benjamins.
- Haiman, John. 1985. *Natural syntax*. Cambridge: Cambridge University Press.
- Itkonen, Esa. 1975. *Concerning the relationship between linguistics and logic*. Distributed by the Indiana University Linguistics Club.
- \_\_\_ 1978. *Grammatical theory and metascience*. Amsterdam: Benjamins.

- \_\_\_ 1983. *Causality in linguistic theory*. London: Croom Helm.
- \_\_\_ 1991. *Universal history of linguistics: India, China, Arabia, Europe*. Amsterdam: Benjamins.
- \_\_\_ 2000. Tolkaappiyam: the basic work of the Ancient Tamil language and culture. *The SKY Journal of Linguistics* 13.
- \_\_\_ 2003. *What is language? A study in the philosophy of linguistics*. University of Turku: Publications in General Linguistics 8.
- \_\_\_ 2011. On the role of induction in typological linguistics. *Papers on typological linguistics*. University of Turku: Publications in General Linguistics 15.
- \_\_\_ 2013. Philosophy of linguistics. Keith Allan (éd.): *The Oxford handbook of the history of linguistics*. Oxford: Oxford University Press.
- \_\_\_ & Anneli Pajunen. 2010. *Empiirisen kielitieteen metodologia* [= 'La méthodologie de la linguistique empirique', en finnois]. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Katz, Jerrold. 1981. *Language and other abstract objects*. Oxford: Blackwell
- Pajunen, Anneli. 2006. Verbisanaston uudistuminen [= 'Le renouvellement du lexique des verbes', en finnois]. *Puhe & Kieli / Speech & Language* 26,4.
- Vainio, Seppo, Jukka Hyönä & Anneli Pajunen. 2003. Facilitatory and inhibitory effects of grammatical agreement: Evidence from readers' eye fixation patterns. *Brain & Language* 85.
- \_\_\_ 2008. Processing modifier – head agreement in reading: Evidence for a delayed effect of agreement. *Memory & Cognition* 2.
- \_\_\_ 2009. Lexical predictability exerts robust effects on fixation duration, but not on initial landing position during reading. *Experimental Psychology* 56,1.