



HAL
open science

Les données en linguistique : choix, nature et exploitation. Une vue personnelle

Francis Cornish

► **To cite this version:**

Francis Cornish. Les données en linguistique : choix, nature et exploitation. Une vue personnelle. Travaux de Linguistique : Revue Internationale de Linguistique Française, 2019, 79 (2), pp.7-42. 10.3917/tl.079.0007 . hal-03768993

HAL Id: hal-03768993

<https://hal.science/hal-03768993>

Submitted on 5 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les données en linguistique : choix, nature et exploitation. Une vue personnelle

Travaux de Linguistique 79, 2019, pp. 7-42

Francis CORNISH*

Résumé

L'article est une réflexion critique sur la matière première des travaux en linguistique, à savoir, les exemples présentés pour l'analyse. La discussion permettra de mettre en évidence l'attitude du linguiste vis-à-vis de la langue et du langage sous-tendant son choix d'exemples et la manière dont il les traite. Deux grands types de données sont contrastés : les exemples inventés par l'auteur lui-même, puis des extraits choisis par l'auteur provenant de corpus d'énoncés produits par des locuteurs ou scripteurs afin de réaliser un but communicatif ou expressif quelconque. Les avantages et inconvénients de chaque type sont dégagés, tant en ce qui concerne les données elles-mêmes que par rapport à l'usage que l'auteur-linguiste en fait pour sa démonstration. L'article examine également les statuts attribués par l'auteur aux exemples qu'il présente pour mener son analyse, et fait valoir que les mêmes statuts ((a)grammatical, (a)sémantique et (in)cohérent) peuvent caractériser les données attestées tout comme les exemples construits.

Mots clés : contextualisation ; corpus d'usages ; exemples inventés ; exemples attestés ; prosodie

Abstract

The article is a critical examination of the raw material of work in Linguistics, namely the examples presented for analysis. The discussion will enable us to point up the linguist's attitude towards language(s) underlying his or her choice of examples and the way in which s/he treats them. Two main types of examples are contrasted: those invented by the author him/herself, and extracts selected by the author from corpora of utterances produced by speakers or writers in order to achieve a communicative or expressive goal of some kind. The advantages and disadvantages of each type are pinpointed, both in terms of the examples themselves and of the use made of them by the author in his or her demonstration. The article also examines the statuses attributed by authors to the examples they present for analysis, and argues that the same statuses ((un)grammatical, (a)semantic and (in)coherent) may characterize attested examples just as well as constructed ones.

Keywords: contextualization; corpora of utterances; invented examples; attested examples; prosody

1. Introduction

Le choix et la nature des données, ainsi que leur utilisation, sont à l'évidence tout à fait centraux dans la démarche linguistique, et ils méritent bien à ce titre une considération particulière : c'est là le but de cet article. En effet, la linguistique est sans doute la seule entreprise scientifique qui a le langage à la fois comme objet et comme moyen de l'étude (ce dernier s'érigeant en métalangage). Par ailleurs, le type de données que l'on présente pour l'étude est bien sûr fonction des objectifs de celle-ci. Dans ce qui suit, je vais me concentrer sur le choix et la nature des données employées surtout dans des études de la référence

* Laboratoire CLLE-ERSS et Université de Toulouse 2 ; Jean-Jaurès, 5, Allée Antonio Machado, 31058 Toulouse Cedex 9 ; cornish@univ-tlse2.fr. Je remercie Anne-Catherine Simon, Anita Fetzer, Georgeta Cislaru, Michel Aurnague, Mohammed Jadir, Josep Ribera et Christopher Butler, ainsi que les trois relecteurs anonymes de la revue, pour leurs remarques sur des versions antérieures de cet article.

discursive, car c'est là le domaine dans lequel j'ai travaillé le plus ces dernières années. Nous verrons en cours de route comment un phénomène linguistique majeur (à savoir, l'anaphore) est encore souvent défini, et ce de façon erronée selon nous, en se fondant sur un seul type de données (les exemples forgés par le linguiste) plutôt que sur un autre (les données naturelles, attestées, surtout orales). Pour distinguer formellement les deux types de données dans le texte, celles relevant du premier seront présentées en caractères romains, et celles du second en italiques. L'aspect central au niveau des données, en dehors de leur choix, est la manière dont le linguiste les présente, traite, et évalue : comme nous le verrons, c'est là surtout que le bât blesse, dans certains cas du moins¹.

L'article est organisé de la façon suivante. La section 2 passe en revue un certain nombre de notions et distinctions de base qui vont nous être utiles dans le corps de l'article. Quant à la section 3, elle porte sur la question des exemples inventés, soulignant surtout leurs inconvénients, mais relevant aussi quelques avantages ponctuels. La section 4 compare les exemples tirés de corpus d'énoncés attestés (au sens large du terme), en mettant en exergue certains problèmes liés à leur exploitation par les linguistes. Les données présentées relèvent de l'écrit normé ainsi que de l'oral et de l'écrit non planifiés². Ensuite, la 5^e section se concentre sur le statut attribué par les auteurs à leurs exemples (qu'ils soient inventés ou attestés), en distinguant les propriétés « grammaticale »³ (ou non), « bien formée sémantiquement » (ou non), et « pragmatiquement adéquate quant au contexte d'emploi envisagé » (en tant qu'énoncés réels ou potentiels) ou non. Comme les auteurs employant ces symboles évaluatifs sont surtout ceux qui exploitent des exemples inventés, cette section va proposer des analyses critiques uniquement sur ce type de données.

2. Quelques préliminaires analytiques

Avant d'entrer dans le vif du sujet, passons en revue quelques notions et distinctions d'ordre méta-descriptif (donc, théoriques) qui vont nous servir dans ce qui suit. D'abord la question tout à fait fondamentale du contexte.

En dehors du seul *co-texte*, il nous faudra reconnaître au moins sept autres aspects du contexte, qui seraient les suivants : le domaine de référence d'un texte donné, le discours déjà construit en amont, le genre d'événement langagier en déroulement, le cadre socio-culturel supposé par le texte, les relations interlocutives existant entre les interactants à chaque moment de l'échange, l'occasion de l'échange, et la situation d'énonciation particulière à l'œuvre. Le plus important de ces aspects est sans doute la situation d'énonciation, qui fonctionne comme point d'ancrage par défaut pour le discours à construire (cf. aussi la démarche adoptée par Gadet et Guerin, 2016 : 4 en construisant et en traitant leur corpus de français parlé non standard). Pour l'utilisateur, le contexte est assujéti à un processus continu de construction et de révision au fur et à mesure que le discours se déroule. C'est à travers l'invocation d'un contexte approprié que l'interprète peut convertir en *discours* son traitement de la séquence connexe d'indices perceptibles qu'est le *texte*.

Le *texte* dans la présente acception constitue la trace perceptible d'un acte d'énonciation, qu'il soit écrit ou oral. En tant que tel, il inclut des aspects paralinguistiques d'un tel acte, ainsi que des traits non-verbaux pertinents sémiotiquement, tels que l'orientation du regard, le pointage et autres mimo-gestes – autrement dit, le texte ne se réduit pas aux seuls éléments verbaux. Le *texte* est essentiellement linéaire, à la différence du

¹ Cf. le titre et le contenu de l'article de Jean Lallot (2007), à savoir « Dis-moi comment tu traites les exemples, je te dirai quel grammairien tu es ». Application à Apollonius Dyscole (Syntaxe) ».

² *Nota Bene* : d'autres types de données, par exemple les données sollicitées, ne seront pas traités ici, faute de place. Par contre, les données expérimentales seront examinées brièvement au cours de la sous-section 4.1.

³ Au sens « bien formée du point de vue de la morphosyntaxe », uniquement.

discours (voir ci-dessous). Ainsi, le rôle du *texte* peut être envisagé comme fournissant une séquence connexe d'indices perceptuels (donc le produit physique d'un acte d'énonciation), produite par le locuteur / scripteur à l'intention de son allocataire / lecteur, pour qu'il infère le *discours* (le contenu négocié) qui pourra être associé à une étendue de texte donné, en liaison avec un contexte pertinent. L'input du seul texte à cet égard sous-détermine donc l'interprétation globale qui en sera faite par la suite (donc, le *discours*, dans cette acception).

Le *discours*, en revanche, est l'interprétation révisable, en évolution constante, d'un événement communicatif, qui est construite mentalement et conjointement par les participants au fur et à mesure que le texte et un contexte pertinent sont perçus et invoqués, respectivement. Cette interprétation « en ligne » sera convertie par la suite en une représentation discursive mentale, capable d'être retenue en mémoire à long terme.

Une autre dichotomie qui va nous être utile est celle entre *phrase* et *énoncé*. Une *phrase*, c'est bien connu, est une unité de l'écrit (donc de *texte*, tel qu'on vient de le définir). Elle peut être simple ou bien complexe, composée de plusieurs propositions syntaxiques. Comme l'ont déjà démontré nombre de linguistes (par ex. Gadet, 1990 : 14 ; Blanche-Benveniste, 1990 : 14), l'oral non planifié, lui, ne manifeste pas de *phrases*. Tout au plus un segment d'oral contient-il des propositions syntaxiques. Les unités constitutives d'un texte oral sont plutôt des unités intonatives minimales, ainsi que des *périodes*. Une *période* est définie par Simon (2004 : 232) comme « [Une] unité intégrative de la clause caractérisée par un intonème conclusif »⁴.

En revanche, un *énoncé* est le produit d'un acte d'énonciation, qui en tant que tel, partage avec la *phrase* la propriété d'être une unité de *texte*. Un énoncé suppose donc un énonciateur, un contexte d'énonciation (un lieu, un moment ou un intervalle où et pendant lesquels il aura été produit), une occasion d'emploi, etc. et (normalement) un destinataire. Une phrase, elle, n'a aucune de ces propriétés. Elle relève uniquement de la métalangue du grammairien-linguiste, et ne porte pas de marques de sa « production » en situation.

Quant à la langue dans son ensemble, celle-ci, dans une optique fonctionnaliste, constitue un « trésor » (pour employer le terme évocateur proposé par F. de Saussure), donc un ensemble structuré de ressources langagières, sédimentées depuis les siècles à partir d'usages faits par ses locuteurs, permettant ainsi à ceux-ci de s'exprimer et de communiquer avec d'autres en situation. Nous distinguons donc entre la langue en tant que système, « trésor » de ressources pour l'usage, existant indépendamment des emplois qui en sont faits, et son exploitation par les usagers pour effectuer des actes de communication à chaque fois uniques.

À présent, penchons-nous sur ce qui peut dysfonctionner dans tel ou tel emploi de la langue (en liaison, bien sûr, avec d'autres capacités de communication). Laissant de côté pour l'instant la dimension purement phonique d'un énoncé, la morphosyntaxe d'une proposition syntaxique peut enfreindre une ou plusieurs règles ou conventions à ce niveau de structure. Par convention, de telles malformations sont marquées dans les travaux des linguistes par un astérisque.

Sur ce symbole, il convient de faire trois remarques. D'abord, celui-ci n'est plus l'apanage exclusif des linguistes opérant dans le giron de la Grammaire Générative chomskyenne (ou dans d'autres modèles de type formel analogues, comme la HPSG)⁵,

⁴ Voir à ce propos également l'analyse de Luzzati (1985) portant sur des extraits de l'émission radiophonique « Le masque et la plume » (*France Inter*). Pour sa part, Luzzati (1985 : 63, n. 1) définit une *période* comme une « unité sémantique et formelle qui comporte trois éléments : la *tension*, la *condition* et la *résolution* ».

⁵ Sampson (2017 : 35-55) remet en cause de façon radicale la notion chomskyenne de « grammaticalité », en citant des exemples célèbres du genre *The boy may frighten sincerity* 'Le garçon peut faire peur à la sincérité' (Sampson, 2017 : 49) qui sont astérisqués par leur auteur, mais qui sont en réalité sémantiquement bizarres, et non grammaticalement mal formés (au plan purement morphosyntaxique). Ils mériteraient donc plutôt le symbole ! (= sémantiquement déviant : voir plus loin) que *.

puisque'il est largement employé dans des cadres souvent très éloignés de cette mouvance, ne partageant pas les mêmes présupposés concernant la nature de la langue et du langage (par exemple, par Givón, 2017, linguiste d'obédience cognitivo-fonctionnaliste, ou Leeman-Bouix, 1994, ouvrage 'grand public' critique sur les jugements des puristes et de la grammaire traditionnelle).

Ensuite, comme nous le verrons dans la section 5, ce symbole est souvent employé abusivement pour caractériser des malformations autres que celles relevant de la stricte morphosyntaxe (par exemple, pour marquer l'incongruité de deux phrases successives n'ayant aucune relation sémantique ou discursive l'une avec l'autre).

Enfin, nous ferons valoir que, contrairement à la Grammaire Générative et des modèles formels analogues, on ne peut pas normalement juger de la bonne formation morphosyntaxique d'une expression langagière sans tenir compte du sens qu'elle sera censée exprimer⁶. Ce type de jugement ne peut donc pas, en principe, être établi indépendamment. Cela va dans le sens des modèles fonctionnalistes du langage, mais est en contradiction avec les approches formelles, qui tendent à envisager la syntaxe comme étant « autonome » par rapport à la sémantique, la phonologie ou la pragmatique. La morphosyntaxe constitue en revanche pour les approches fonctionnalistes l'enveloppe expressive du langage (cf. Dik, 1997 ; Hengeveld et Mackenzie, 2008, *inter alia*). De même, les constructions syntaxiques véhiculent un sens, elles ne sont donc pas sémantiquement neutres (des illustrations s'en trouvent dans les exemples [1a] (§ 3), avec la construction dite à « montée du sujet », puis la seconde phrase de [2a] (§ 3) et [6] (§ 4), avec le mode déclaratif).

Pour ce qui concerne les unités lexicales, cependant, de par leur interprétation en contexte d'emploi, celles-ci se trouvent à l'interface entre la langue en tant que système et son emploi effectif en situation. En effet, la signification conventionnelle (en langue, donc) de ces unités est toujours modulée (enrichie ou restreinte) en fonction du contexte qui préside à l'énonciation. Quant à la mise en relation de la signification des éléments lexicaux dans une expression, celle-ci peut également donner lieu à une impasse (ici sémantique), de par l'incompatibilité des significations de lexèmes figurant dans un même cadre syntaxique (une même construction, par exemple). Cependant, au niveau co-énonciatif, le sujet interprétant pourra bien, selon le contexte et son sentiment de la manière dont le locuteur ou scripteur est en train d'employer l'expression contenant des éléments signifiants *a priori* incongrus, leur donner une interprétation étendue, figurée (synecdoque, personnification, de type oxymore, ou autre). Mais l'input d'une telle réinterprétation pourra tout de même selon le cas être l'incompatibilité des sèmes concernés au premier degré, au niveau littéral. Une telle configuration sera marquée par le point d'exclamation (voir la note 5 pour une illustration).

Enfin, au niveau énonciatif, pragmatico-discursif, une expression parfaitement bien formée au niveau de la langue-système, pourra ne pas convenir en tant qu'énoncé au contexte dans lequel elle est ou pourra être utilisée. L'aspect du contexte en question pourrait, par exemple, être celui relevant du *genre* qui aura été invoqué précédemment, dans lequel l'expression ne pourra pas être aisément intégrée (car relevant d'un genre distinct). Le symbole employé pour marquer ce statut est le dièse. En principe, des déviations au niveau du (co-)texte (incohésion textuelle) ainsi que du (co-)discours (incohérence discursive) relèveraient aussi du symbole # ; mais pour distinguer chacun de ces types, on se propose de lui ajouter des diacritiques en exposant : ^{'CST'} pour les cas d'incohésion textuelle, et ^{'CRD'} pour ceux d'incohérence discursive.

Voir surtout la section 5 pour la question centrale des statuts attribués en fonction de ces différents niveaux d'analyse.

⁶ Mis à part les cas, pour ce qui concerne les modèles formels, où il y a des quantifieurs dans une phrase ayant des portées logiques différentes, qu'il faut donc distinguer, et ceux où la co-indiciation permet de différencier deux marqueurs référentiels indiquant telle ou telle relation de coréférence visée.

3. Les exemples inventés

Une caractéristique majoritairement adoptée en linguistique de nos jours s’agissant des données sur lesquelles doit porter l’analyse est celle d’être attestée par l’usage plutôt que forgée par le linguiste lui-même.

Les inconvénients de cette dernière approche sont connus : il peut y avoir circularité dans la mesure où le linguiste (même inconsciemment) fait en sorte que sa théorie ou son hypothèse quant au fonctionnement de tel ou tel aspect de la langue ou du discours se reflète dans les données mêmes forgées pour la démonstration. Hanks (2013 : 20), cité par Sampson (2017 : 92), opère une distinction importante à cet égard :

“There is a huge difference between consulting one’s intuitions to explain data and consulting one’s intuitions to invent data. Every scientist engages in introspection to explain data. No reputable scientist (outside linguistics) invents data in order to explain it. It used to be thought that linguistics is special — that an exception could be made in the case of linguistics — but comparing the examples invented by linguists with the actual usage found in corpora shows that this is not justifiable.”

De même, les données construites revêtent parfois un caractère peu naturel, rendant difficile sinon impossible l’invocation d’un contexte d’usage potentiel qui permettrait de les motiver, et courent le risque de se rapprocher de l’écrit normé (ceci en fonction des caractéristiques du linguiste typique : éducation classique fondée sur l’écrit littéraire, sentiment aigu de l’importance de la « norme » qui régit la correction des données langagières...).⁷ Deux exemples anglais illustrant cette tendance sont présentés en [1] :

- [1] a. The men believe pictures of themselves to amuse me.
 ‘Les hommes croient que des images (ou photos) d’eux-mêmes m’amusent’
 (Barss, 2003, ex. 16 : 7)
- b. Joe wonders which book that Mary checked out of the library she put where.
 (Lit.) ‘Joe se demande quel livre emprunté par Mary à la bibliothèque elle a mis où’ = ‘Joe se demande où Mary avait mis quel livre qu’elle avait emprunté à la bibliothèque’ (Barss, 2003, ex. 17 : 7)

Ces exemples semblent relever d’une simple combinatoire (grammaticalement acceptable, sans doute) de mots, sans véritable prise sur une situation d’énonciation quelconque. Ils sont de plus dépendants par rapport à la théorie linguistique que l’on essaie de justifier : car c’est la théorie qui est prioritaire, les exemples en découlant directement (cf. aussi Willems, 1985 : 88 à ce propos). Il serait d’ailleurs surprenant que ces types de phrases soient représentés dans un quelconque corpus d’usages réels (voir la section 4). Chafe (1994), précisément, adopte un symbole pour marquer de tels exemples (forgés afin de justifier un argument théorique, mais d’une probabilité faible d’être employés par un usager réel pour réaliser un but communicatif authentique) : à savoir, le symbole ‘~’. Schütze (à paraître) présente bon nombre d’autres exemples du même acabit, tirés des travaux de linguistes générativistes, et qui avaient été proposés pour évaluation à des sujets natifs non-linguistes. Parmi les résultats obtenus, certains n’avaient pas du tout été ceux escomptés, car manifestement, les sujets avaient eu bien du mal à les traiter, analyser et interpréter comme

⁷ Voir Linell (2005) pour une démonstration complète et systématique de la manière dont l’écrit a pris le dessus sur l’oral dans l’histoire de la linguistique concernant les données utilisées, et les conséquences qu’a eues cette préférence.

cela avait été prévu⁸. Schütze (2016 : 13), à juste titre, qualifie cette activité d'évaluation de la grammaticalité de phrases inventées de "metalinguistic performance". Celle-ci consiste en une forme d'introspection, « objectivant » un segment de texte en l'abstrayant (dans la mesure du possible) des « confounds » éventuels (cf. la note 8), afin d'en cerner le degré de grammaticalité à l'œuvre. Voir son schéma « Model of Language Processing », Figure 6.1, p. 172, qui présente l'ensemble des connexions qui existeraient entre les différents facteurs en jeu ici. La grammaticalité ne doit pas être confondue, selon Schütze, avec l'analyse grammaticale (le « parsing », cf. p. 160-161) qui est impliquée dans le traitement normal du texte en emploi réel. Un segment de texte présentant une difficulté d'analyse ne signifie pas pour autant que celui-ci soit agrammatical.

Penchons-nous d'un peu plus près sur les exemples en [1], tous deux peu idiomatiques. Dans le cas de [1a], cela est dû selon nous à la construction employée (...V [SN_[indéf] V_{inf} SN]), avec le SN indéfini inanimé *pictures of themselves* à la fois objet du verbe principal *believe* et sujet du verbe enchâssé à l'infinitif *amuse*. Une forme alternative de cette construction, avec un SN animé, serait bien plus attendue : par exemple (*The men*) *believe John to know the truth of the matter* ' (Les hommes) croient que John connaît la vérité sur l'affaire'. Cette construction anglaise à « montée » du sujet de la complétive⁹ n'existe pas en français avec des verbes de type *croire*, d'où la traduction de cet exemple avec emploi d'une complétive finie. Cette dernière forme syntaxique aurait d'ailleurs été plus naturelle en anglais ici aussi, à savoir : *The men believe that pictures of themselves amuse me*. La version en [1a] avec le verbe enchâssé à l'infinitif et son sujet inanimé indéfini, sans subordonnant manifeste, présente le référent du sujet « monté » de la proposition infinitive comme étant agentif ou du moins contrôlant, exerçant ainsi une influence sur la mise en place de l'état de choses 'my being amused', 'le fait que je sois amusé', ce qui est *a priori* en contradiction avec le statut d'inanimé de ce référent. D'où donc la gêne que l'on peut ressentir devant cet exemple, pourtant non étoilé, ni même interrogé.

Quant à l'exemple [1b], contenant deux éléments en *wh-* (*which* et *where*), celui-ci semble encore plus bizarre. Une version bien plus naturelle serait : *Joe wonders where Mary put which book that she checked out of the library* 'Joe se demande où Mary avait mis quel livre qu'elle avait emprunté à la bibliothèque'. Du point de vue énonciatif, cette phrase pourrait servir à exprimer une « question écho » indirecte. Le facteur principal derrière la lourdeur de la réalisation dans [1b] est le fait que l'objet direct du verbe *wonders* se voie étendu au moyen d'une relative (*that [...] library*), ce SN complexe précédant mais non suivant la proposition interrogative indirecte plus courte et moins complexe *she put where*. La version ci-dessus où le SN complexe contenant la relative est placé à la fin de la phrase serait de ce fait plus acceptable, attendue. En outre, la suppression de la relative rendrait la phrase encore plus recevable - mais bien entendu, cela aurait empêché l'auteur de mener à bien son projet de démonstration du point théorique à l'œuvre ici ! Selon nos analyses, l'exemple [1a] aurait dû être marqué par le symbole '!', car il y a une contradiction évidente entre l'apport sémantique de la construction infinitive à « montée du sujet » et la nature du référent de ce sujet ; et [1b], lui, par le symbole '#^{CST}', puisque la textualisation qui lui correspond sera de nature à compliquer la saisie par l'interprète de sa structure informationnelle (cf. le § 2 *supra*).

⁸ Schütze suggère parmi les moyens à envisager pour éviter ces « confounds », comme il les appelle, dans des enquêtes futures, celui consistant à « fournir de meilleurs conseils sur la manière dont il conviendrait de lire les phrases proposées, par exemple, en les insérant dans un contexte » (Schütze, à par. : 26). De plus, écrit-il, « il devrait être évident qu'un « copié-collé » direct des exemples tirés d'articles de linguistique pour créer des stimuli expérimentaux va en général donner lieu à des résultats ininterprétables ». Étonnantes admissions de la part d'un générativiste ! Mais bienvenues, tout de même.

⁹ Pour employer ici un terme plus traditionnel, on pourrait qualifier cette construction d'« accusatif avec infinitif ».

Les exemples en [1] (de même que ceux proposés aux sujets non-linguistes relevés dans Schütze, à paraître) montrent d'une manière exemplaire que pour leur auteur, ce n'est que le plan de l'expression langagière qui compte¹⁰, toute autre considération - invocation d'une situation de communication, d'un locuteur et d'un destinataire (un interlocuteur potentiel), d'une occasion de l'emploi, en somme d'un contexte dans lequel une expression peut être située et utilisée - étant étrangère à la démarche. Selon cette conception, c'est donc l'expression et l'expression seule, qui fait l'objet de la linguistique. Les données présentées sont ainsi de simples *phrases* et non des *énoncés* potentiels. Or, la langue est plus qu'un code abstrait, elle sert essentiellement de moyen de communication entre des humains dans des interactions psycho-sociales potentielles, et ses propriétés intrinsèques découlent de cette fonctionnalité. Il s'ensuit qu'on ne peut pas en faire abstraction en l'étudiant scientifiquement.

L'absence quasi totale de contexte fourni constitue un inconvénient majeur d'une trop grande dépendance par rapport aux données construites. Car les données construites par le linguiste doivent de ce fait s'interpréter en fonction d'un « contexte par défaut » - autrement dit, d'un contexte « idéalisé », en quelque sorte. Ce contexte dérive donc entièrement du texte de l'exemple lui-même, qui revêt de ce fait un caractère immanent (comme cela est à l'évidence le cas des deux exemples inventés déjà vus en [1]). Il est intéressant de ce point de vue que selon Bader et Häussler (2010 : 320), lors de la présentation sans contexte à des sujets natifs de certaines phrases allemandes marquées, où le sujet suivait et non précédait l'objet après le verbe fini, ceux-là avaient tendance à les juger comme agrammaticales ; or, quand un contexte était proposé où un focus étroit était imposé sur le référent du sujet, elles étaient jugées grammaticales et acceptables (car dans ce cas, l'ordre non canonique du sujet et de l'objet était contextuellement motivé). À l'évidence, donc, dans le premier cas, les sujets concernés avaient évalué l'exemple en question en tant que *phrase*, alors que dans le second, où un contexte motivant était fourni, c'est en tant qu'*énoncé* qu'ils l'avaient jugé (Schütze, 2016 : 149 remarque précisément ce même facteur en invoquant la capacité d'un contexte linguistique à motiver l'acceptabilité de constructions marquées, telles que les clivées en anglais. Voir Schütze, 2016 : 20-21 pour la distinction entre « acceptabilité » et « grammaticalité »).

Givón (2017) éclaire la question de l'emploi quasi exclusif des exemples inventés, notamment dans les travaux conduits dans la mouvance générative, en contrastant l'oral non planifié et l'écrit soutenu, normé. Il rapproche la « compétence » langagière chomskyenne de l'essentialisme de Platon ainsi que de la préférence chez les philosophes du langage (par exemple, Russell et Carnap) pour le langage formel, logique (Givón, 2017 : 135). À l'instar de Linell (2005 ; voir la note 7), Givón prétend, en compagnie de bon nombre d'autres linguistes – sans doute la majorité aujourd'hui –, que le langage se manifeste dans sa forme première à travers l'oral non-planifié, en situation d'échange face-à-face ; l'écrit, lui, en est une manifestation secondaire, qui se greffe pour une large part sur l'oral, mais s'en éloigne par d'autres aspects. Il existe d'ailleurs des langues, aujourd'hui disparues ou bien présentes actuellement, pour lesquelles il n'y a aucune forme écrite du tout.

Or, il y a une « grammaire de l'oral » qui est bien distincte de la « grammaire de l'écrit » (cf. aussi Gadet, 1990 : 14-15 pour ce qui concerne les faits du français) : l'oral non planifié n'est absolument pas une version « imparfaite » de l'écrit normé¹¹, assimilable à la

¹⁰ Expression contrôlée uniquement de par son intuition de locuteur natif, qui, mêlant connaissances de la langue puis respect inconscient de la norme régissant l'écrit, n'est pas toujours fiable, comme le montre souvent la mise en regard de données provenant des corpus (voir aussi Willems, 1985 : 92). Or, Schütze (2016 : 112-114) recommande spécifiquement que ce soient les informateurs naïfs par rapport à la linguistique (moyennant l'imposition d'un certain nombre de conditions afin d'éviter les jugements non pertinents), et non les auteurs-linguistes eux-mêmes, qui devraient évaluer la grammaticalité des exemples proposés.

¹¹ Auquel correspond pour l'essentiel la « compétence », dans la conception de Chomsky.

« performance » chomskyenne, mais possède en propre bon nombre de propriétés distinctives à cet égard : voir la liste de propriétés spécifiques donnée à la p. 132. Voir aussi à ce propos les exemples de cette nature présentés en [8a,b] et [9a-c] au § 4.1 *infra*.

Cependant, il ne s'agirait pas de « bannir » complètement l'emploi des exemples inventés (même si on le pouvait, ce qui est loin d'être certain !), mais de le restreindre nettement, en les employant comme cadres pour vérifier en quelque sorte la valeur des données attestées – qui doivent rester la source primaire de l'analyse linguistique. Car de tels exemples peuvent être très utiles (dans la mesure du possible) dans la modification¹² de tel aspect des données naturelles selon tel ou tel paramètre – par exemple, afin de faire ressortir par contraste ou comparaison une propriété de ces dernières que l'analyste veut y déceler. Autre utilité potentielle des exemples inventés, celle de présenter dès le début d'une étude, de manière générale, le phénomène à cerner. Il est vrai qu'un exemple attesté pourrait également faire l'affaire dans ce cadre ; mais plutôt que de faire une recherche étendue pour trouver un exemple prototypique du phénomène objet de l'étude qui ne revête pas d'aspects strictement non pertinents pour son objectif, un exemple forgé pour la circonstance pourrait plus aisément remplir cette fonction.

Par ailleurs, les exemples forgés sont bien entendu utiles, voire nécessaires, pour la psycholinguistique expérimentale, car ils permettent de contrôler les variables en jeu et par suite de mesurer des comportements liés au processus de compréhension - mais on doit néanmoins avoir toujours à l'esprit leurs inconvénients notoires. Voir à ce sujet la discussion à la fin du § 4.1.

À présent, si l'on s'en tient aux travaux portant sur l'anaphore, il est fréquent que les linguistes d'obédience formaliste-textualiste emploient comme base de leurs analyses des exemples qui consistent en une séquence de deux phrases inventées. La première contiendra l'antécédent textuel (le plus souvent un SN dénotant un être humain), et la seconde un anaphorique qui peut s'interpréter comme entrant dans une relation de reprise (habituellement en termes de coréférence) vis-à-vis de cet antécédent, ou bien comme ayant une référence « disjointe » par rapport à celui-ci. Un exemple bien connu est [2a], proposé par Milner (1982 : 21) :

- [2] a. Ton frère est arrivé hier ; l'époux de Jeanne avait manqué tous ses tirs.
 b. Ton frère est arrivé hier ; il avait manqué tous ses tirs. (ex. [2a] modifié — FC)

L'exemple [2a] (inventé) suscite chez l'interprète un sentiment de gêne analogue à celui que peuvent provoquer les exemples inventés anglais en [1] ci-dessus. Car ici, la phrase initiale sous-entend un segment de discours direct (un « énoncé » potentiel, donc, qui serait produit en situation par un locuteur qui connaît son interlocuteur – à témoin, le tutoiement évident dans *ton frère*). Cependant, la seconde phrase semble être une observation (typique de l'écrit soutenu) provenant d'une tierce personne : cela provient en partie de l'emploi d'un SN défini en position sujet, sans doute à intention de reprise du référent du sujet de la première, le caractérisant en fonction de sa relation maritale avec une dénommée « Jeanne », plutôt qu'en y co-référant au moyen d'un simple pronom de 3^e personne au masculin singulier (comparer la version modifiée en [2b]). C'est cela qui fait que cette suite de deux phrases (telle qu'envisagée en tant qu'input à une unité de discours potentielle) ne serait pas pleinement cohérente à ce titre, car elle manifesterait, tout comme l'exemple [11] au § 5 *infra*, une sorte de « dissonance polyphonique », où les « fils perspectivaux » se trouveraient croisés. De toute

¹² Une fois un exemple attesté modifié, il cesse par-là même d'être « attesté », authentique. Mais une telle manipulation peut bien avoir une utilité heuristique, en comparant le résultat avec la forme ou la signification de l'énoncé d'origine.

manière, au niveau d'une énonciation possible, on ne voit pas pourquoi un usager emploierait un SN défini relationnel pour renvoyer à une entité ('ton frère') qu'il est censée, de par l'emploi de cette dernière appellation, connaître déjà et qui vient tout juste d'être introduite dans ce discours (cela vaut également pour son allocataire ou lecteur, bien sûr). Ce serait là une violation de la seconde maxime de Quantité de Grice (1975) : « Ne faites pas en sorte que votre contribution soit plus informative qu'il n'est nécessaire ».

Ajoutons que le marqueur censé assurer la coréférence dans la seconde phrase de [2a] (le SN défini relationnel *L'époux de Jeanne*) est une expression référentielle quasi autonome, statut voulu précisément ici par Milner (1982 : 21), et non référentiellement dépendante, comme l'est le pronom de 3^e personne en [2b]. À ce titre, il pourrait servir à réaliser l'introduction dans un discours de l'entité qu'il dénote (il constituerait ainsi un « antécédent » potentiel dans la conception classique de l'anaphore). En [2a], donc, il y aurait « coréférence » mais non « anaphore » au sens strict. Du point de vue de la prosodie avec laquelle cette suite de phrases serait prononcée, si ce SN défini est employé dans une fonction de reprise, alors il serait réalisé avec un ton bas, à accentuation faible, sur les deux éléments lexicaux qui le composent, avec une légère remontée vers la fin. S'il est employé pour introduire une entité nouvelle, en revanche, ces deux éléments seraient accentués et à hauteur élevée, à intonation descendante. Ce sera donc uniquement la prosodie qui permettra de distinguer les deux fonctions discursives potentielles ici.

Or, c'est précisément sur la base de tels exemples inventés, caractéristiques de l'écrit normé, que la notion d'*anaphore* est le plus souvent définie. Voici les définitions qu'en proposent Milner (1982 : 18, 20) ([3a]), et Huang (2000 : 1) ([3b]) :

- [3] a. « Il y a relation d'anaphore entre deux unités A et B quand l'interprétation de B dépend crucialement de l'existence de A, au point qu'on peut dire que l'unité B n'est interprétable que dans la mesure où elle reprend — entièrement ou partiellement — A. »... « L'anaphore est directement interprétable à partir du seul contexte linguistique, sans informations touchant les segments désignés [...] ».
- b. « [The term 'anaphora'] is commonly used [in contemporary linguistics] to refer to a relation between two linguistic elements, wherein the interpretation of one (called an anaphor) is in some way determined by the interpretation of the other (called an antecedent) ».

Notons tout de suite que ces deux définitions présentent l'anaphore comme une relation existant uniquement entre segments du *co-texte*, à l'exclusion de toute autre manifestation du contexte, telles qu'énumérées au § 2 *supra*. De telles définitions, qui sont assez répandues dans la littérature, prennent appui sur des données textuelles caractéristiques de l'écrit soutenu, planifié, respectant la norme langagière en vigueur – norme qui précisément impose à tout élément à vocation anaphorique une mise en relation avec un antécédent textuel explicite¹³ se trouvant à proximité et qui va à ce titre lui transmettre son sens et sa référence. Mais cela ne va pas toujours être le cas, même dans l'écrit soutenu ; et d'autant moins dans l'oral non planifié, qui, de par son caractère situé dans un contexte d'énonciation socio-psychologique partagé par les participants, pourra permettre plus aisément de laisser bon nombre de précisions implicites.

¹³ L'écrit planifié, justement, à la différence de l'oral non planifié, va quasiment par définition fournir de telles expressions explicites. Cependant, les textes réels ne satisfont pas toujours à cette condition (voir la note 20 au § 4.1 *infra* pour quelques exemples types).

De toute manière, le marqueur anaphorique (entre autres, formes nulles, pronoms de 3^e personne ou démonstratifs, SN définis ou démonstratifs) possède en propre une capacité de référence plus ou moins indépendante par rapport à celle dont son « antécédent » (là où il existe) sera doté, et, dans une langue à genre grammatical comme le français, peut varier en valeur de genre ainsi que de nombre par rapport à celui-ci (voir Reichler-Béguelin, 1993 ; Cornish, 1987 et Cornish et Salazar Orvig, 2016 : 58-60, 66-70 à cet égard). C'est la prédication indexicale (le segment « hôte ») dans son ensemble, unité contenant le marqueur indexical, qui déterminera l'unité contextuelle de discours qu'elle va cibler, puis permettra d'effectuer une intégration grâce à une ou même deux relations de cohérence. Elle le fera en filtrant, triant les unités candidates à ce choix, et en ciblant l'une d'entre elles qui va constituer le choix préféré. La relation entre « antécédent », à présent situé au niveau du discours et non plus du seul texte, et marqueur anaphorique (ou « indexical », plus généralement) est de ce fait complètement inversée. Voir Cornish (1996) pour une argumentation et une illustration plus amples de ce point.

Notons à cet égard que les « antécédents » textuels ne se signalent pas en tant que tels lorsqu'ils apparaissent dans le fil du texte. Par ailleurs, des anaphoriques peuvent bien être employés pour renvoyer à une entité présumée exister dans un discours en déroulement, sans qu'il y ait d'antécédent textuel à proximité dans le co-texte pour permettre sa « résolution »¹⁴ : ces « entités de discours » peuvent entrer dans un discours par le fait d'être mutuellement perçues et focalisées par les participants dans la situation d'énonciation, ou bien par inférence. Voir Cornish et Salazar Orvig (2016 : 62-66) pour une présentation plus complète de cette conception de la référence indexicale discursive ; puis Smith *et alii* (2005) pour des exemples et analyses de la manière dont les locuteurs d'anglais américain introduisent des référents dans un discours. Ce sont précisément les exemples attestés, fruits d'un emploi par un locuteur ou scripteur pour effectuer un acte de communication, qui permettent cette nouvelle perspective. C'est donc vers ce domaine que nous nous tournons à présent.

4. Les données provenant des corpus d'énoncés attestés : le problème de la décontextualisation

La constitution des corpus d'énoncés de types variés et le recours à ces sources de données sont très répandus à l'heure actuelle en linguistique¹⁵. Les données ainsi repérées permettent en principe une vision beaucoup plus fine et réaliste du fonctionnement de la langue et de ses emplois potentiels que ne peut offrir la démarche traditionnelle avec ses exemples inventés, notamment en fournissant des gages statistiquement validables des phénomènes étudiés.

Cependant dans des approches visant à cerner des phénomènes d'ordre discursif et fonctionnel, l'intérêt d'un tel accès peut s'avérer limité¹⁶. En prenant comme exemple mes propres recherches sur la référence indexicale en discours, la dimension qualitative

¹⁴ Nous en verrons une illustration dans l'exemple [6] plus loin (§ 4), extrait du tout début d'un texte écrit littéraire. Toutefois, selon la définition de Milner en [3a] ci-dessus, un tel cas de figure devrait être exclu. Il en va de même des exemples d'oral et d'écrit non planifiés [8a-c] et [9a-c] au § 4.1 *infra*. Voir aussi Gerrig *et al.* (2011) et Kitzinger *et al.* (2012) pour d'autres illustrations à l'aide d'extraits de corpus oraux.

¹⁵ Notamment, mais pas exclusivement, parmi les approches non formelles ; certes, il y a bien des approches formelles qui se prémunissent de données provenant de corpus. Par exemple Lee (2017), cité plus bas. Cependant, dans ce cas, c'est généralement de corpus d'énoncés écrits qu'il s'agit, comme dans son cas.

¹⁶ The use of fully automatic methods is generally impossible in corpus-based syntactic or discourse-oriented contrastive research [...] De même, Cislaru et Sitri (2009 : 41-2) relèvent que l'analyse automatique seule n'est guère suffisante pour l'analyse du texte en tant qu'unité linguistique. Les auteures plaident pour une complémentarité d'analyses quantitative et qualitative dans ce cadre.

(évaluation de la nature de la référence d'une expression quelconque, en fonction de son contexte) y prime sur le purement quantitatif. De ce fait, il est difficile, voire impossible, d'obtenir des données de chaque type ou sous-type par le biais d'instructions opérées sur tel ou tel corpus de textes. Toutefois, les analyses de corpus modernes peuvent bien se prévaloir de nombre de corpus actuellement disponibles qui sont annotés de façon à capter des faits se situant à des niveaux plus élevés que la seule morphosyntaxe —même si une grande partie de ces annotations doit encore être effectuée manuellement.

Les instructions permettant l'accès aux données fournies par les corpus se caractérisent en fonction d'une requête de formes textuelles occupant telle ou telle position dans le flux du co-texte. Les résultats d'une telle recherche devront par la suite être soumis à une analyse qualitative au cas par cas – et encore, nombre des formes relevées s'avèreront fatalement non pertinentes pour les besoins de l'étude. Bien entendu, l'obtention de données par ce biais est tout à fait utile pour ce qui concerne les études de phonologie, de morphologie ou de morphosyntaxe, ou même de faits de langue en tant que tels : par exemple, l'établissement dans les textes des degrés de fréquence d'occurrence relatifs des membres d'une micro-opposition quelconque, ce afin de déterminer lequel en serait le membre marqué et lequel le membre non-marqué. C'est même là la seule méthode empirique possible pour le faire.

L'utilité des corpus pour l'étude des phénomènes discursifs consiste plutôt dans le fait qu'ils offrent à l'analyste des pans de textes réels, de genres et registres divers, oraux comme écrits (cf. aussi Greidanus et Callebaut, 1990 : 83). Ces textes auront été pris en charge et produits par des locuteurs et interlocuteurs pour accomplir des buts communicatifs personnels.

Toutefois, l'exploitation de données attestées issues d'un corpus peut achopper à un inconvénient important : si dans un travail portant sur la référence ou sur un autre aspect du discours, on présente les extraits d'un corpus quelconque directement pour analyse, sans autre indication que leur seule provenance (ce qui constitue déjà un début de re-contextualisation), alors ils pourront revêtir un caractère artificiel – car inévitablement amputés de leur contexte d'origine. À ce titre, quoiqu'attestés, ils se rapprochent des données inventées des « armchair linguists » ; car dans ce cas (un comble !), tout comme avec les données construites, l'unique moyen à disposition du destinataire de l'étude en question serait d'essayer de les re-contextualiser en fonction du seul co-texte à l'œuvre, ce qui ne permettrait qu'avec difficulté d'avoir accès au discours qui pourra être associé à ces fragments textuels.

Un exemple illustrant ce cas de figure provient d'un article récent portant sur le fonctionnement en discours de l'adverbe temporel anglais *now* (cf. Lee, 2017 : 630, ex. (22)). Dans l'extrait [4], tiré de façon « anonyme » du corpus BNC, l'auteure écrit en caractères gras les mots « rode » et « regretted now », ce afin de signaler que c'est en fonction du moment désigné par le temps du verbe *rode* (le prétérit) que l'adverbe *now* est interprété. Or, mon but dans l'analyse de cet exemple est plus général, et ne concerne pas uniquement l'interprétation de l'adverbe temporel *now* :

- [4] *Tamar had no alternative but to do as he said, and they **rode** from the yard without another word. How bitterly she **regretted now** that she had not revealed the episode with Davis before her marriage to Stephen* (aucune indication quant à la source exacte de cet extrait n'est fournie) 'Tamar n'avait d'autre choix que de faire ce qu'il lui avait dit, et ils montèrent à cheval (ou à vélo ?) depuis la cour sans mot dire. Comme elle regrettait amèrement maintenant qu'elle n'avait pas révélé l'épisode avec Davis avant son mariage avec Stephen !'

Ici, il s'agit non pas de déterminer la bonne formation de l'extrait en tant que **forme** (donc,

comme fragment de **texte**) exprimant un sens quelconque, mais d'évaluer le **discours** pouvant être associé à cette séquence. Cela suppose bien entendu la prise en compte obligatoire d'un contexte d'emploi (ici, nécessairement inféré par le lecteur de cet extrait), afin précisément de créer du discours : voir le § 2 *supra*. À ce titre, l'on remarque dans l'extrait [4] bon nombre d'éléments supposant un contexte particulier pour leur interprétation, obligeant le lecteur de ce fait à opérer un certain nombre d'inférences pour ce faire, et surtout pour établir les liens supposés entre les situations et individus évoqués : la référence des noms propres (*Tamar, Davis, Stephen*), les relations existant entre ces personnes, les SN définis (*the yard* – quelle « cour » ?, *the episode with Davis, her marriage to Stephen*), les ellipses (*[...] to do as he said* – qu'aurait-il dit ?), ou le fonctionnement des pronoms et du déterminant possessif (*he, she* (trois occurrences), *they, her*). La référence de *they* à la première ligne englobe par défaut sans doute 'Tamar' et celle supposée de *he* (mais cette interprétation n'est aucunement la seule possible ici) ; le pronom féminin *she* et le déterminant possessif *her* (lignes 2 et 3, respectivement) renvoient sans doute à 'Tamar', mais tel quel, tout lecteur ne comprendra pas nécessairement que ce nom « étranger » renvoie à une personne de sexe féminin.

De toute façon, ce sera la relation anaphorique induite chez le lecteur qui l'établira indirectement. Lee propose (2017 : 630) que le lecteur inférera le plus naturellement que 'Tamar aura eu un sentiment de regret en montant à cheval [« *rode* » : mais ce verbe peut également être compris comme renvoyant au transport en vélo, en moto...] depuis la cour'. Elle omet toutefois de remarquer que cette inférence en suppose deux autres, plus 'locales' : à savoir, que l'une des personnes évoquées par *they* est 'Tamar', et que, comme nous l'avons déjà suggéré, celle-ci ou celui-ci fait donc partie du groupe d'individus que ce pronom désigne.

En fait, le sens contextuel de *rode* est difficile à cerner exactement : ces personnes sont-elles montées ensemble ou séparément (ce serait certainement le cas, par défaut, si le mode de transport était le vélo, ou le cheval) ? Mais la précision « without another word » suggère que chacun est monté sur son propre cheval, vélo etc. Il est vrai que l'interprétation de l'adverbe *now* pose (ostensiblement, du moins ; voir plus bas) moins de problèmes immédiats ; mais *rode* ne peut être contextualisé en l'absence d'informations supplémentaires. Ensuite, la référence exacte du pronom masculin *he* à la ligne 1 n'est pas évidente. À ce sujet, on pourrait penser qu'il s'agirait du dénommé « Davis », car selon ce qui est affirmé de ce personnage (« Tamar had no alternative but to do as *he* said »), il semble clair que l'homme concerné avait exercé une contrainte quelconque sur Tamar ; et de plus, que selon la seconde phrase on apprend que Tamar « regrettait de n'avoir pas révélé l'épisode avec Davis avant d'épouser Stephen ». On peut donc à juste titre conclure, provisoirement du moins, que c'est ce Davis qui avait fait pression sur Tamar pour qu'elle révèle cet « épisode » à Stephen avant son mariage avec lui. Lee (2017: 630) soutient bien que « the event in the preceding clause [certainement l'initiale ici] provides an explanation for the event in the *now* clause ». Mais elle ne dit pas en quoi cela en serait une.

À l'évidence, on est en présence ici d'une instance de « pensée et perception représentées, subjectives » (cf. la notion de « pensée indirecte libre » : Blakemore, 2010, *inter alia*) : ce sont là donc les pensées mêmes de Tamar, qui les assume en tant que sujet de conscience, et non (directement) celles émanant du narrateur. À ce titre, on pourra proposer l'hypothèse selon laquelle dans la seconde phrase, l'adverbe *now*, à fonction déictique, n'est pas relié simplement au repère temporel externe que constitue la référence du prétérit dont est marqué le verbe fini *rode* (comme l'affirme Lee), mais renvoie en réalité au niveau interne, à l'idée que Tamar viendrait de former : à savoir 'qu'elle se sentait incapable de faire autre chose que ce qu'il (donc par hypothèse, 'Davis') lui avait enjoint de faire'. C'est cette pensée-là, qui, selon toute probabilité, l'aura amenée à regretter à présent de n'avoir pas révélé (à Stephen, sans doute) l'épisode avec Davis avant son mariage avec celui-là. La situation est

donc autrement plus complexe que ne le laisse entendre Lee (2017). On voit ici, donc, à quel point le lecteur est contraint de tirer nombre d'inférences en lisant cet extrait, afin d'opérer les connexions nécessaires pour une pleine interprétation de ce fragment, présenté totalement hors contexte.

Cette situation rappelle l'expérience qu'on peut avoir de visionner un film ou d'écouter une émission de radio dont on aura manqué le début : pour en saisir la thématique, il faudra dans ce cas faire rapidement et rétroactivement un ensemble d'inférences ou d'hypothèses à partir de la scène ou du passage que l'on voit ou entend, afin de retrouver ou de reconstituer le fil de l'intrigue filmique ou du contenu de l'émission radiophonique (une forme d'« abduction », donc). L'exemple le plus extrême de ce cas de figure est la situation – rare, toutefois, mais qui existe bel et bien – où un présentateur de journal à la radio annonce un reportage sur telle ou telle thématique, mais où le technicien aux manettes déroule à sa place un autre enregistrement, prévu pour plus tard dans le journal en question. En voici un exemple authentique :

- [5] [Contexte annoncé par le présentateur d'un journal radiophonique : quelqu'un essaie d'atteindre le Pôle sud tout seul et à pied, sans aide extérieure]
Début du reportage pré-enregistré : *One of my staff discovered signs of life [...]*
'L'un(e) de mes employés a découvert des signes de vie...'

Ce fragment ne faisait pas partie du reportage pré-enregistré prévu pour l'item d'informations qui venait d'être annoncé. Cependant, même avant que l'erreur fût reconnue par le présentateur, l'influence déterminante du contexte était si forte que (en compagnie de bien d'autres auditeurs sans doute), j'avais effectué rapidement quasi automatiquement sur le champ plusieurs inférences afin d'essayer d'intégrer de façon cohérente l'information fournie par ce fragment de texte avec le contexte indiqué : à savoir, que l'individu dont il s'agissait n'avait pas pris contact par radio avec les services d'urgences depuis un certain temps, qu'en conséquence, on avait envoyé une équipe de secours pour venir à sa rescousse, et que l'un(e) des « employés » du locuteur interviewé (certainement le directeur de l'équipe en question) avait remarqué sur place de faibles trémoussements dans la neige (« des signes de vie »)...

Une autre illustration – très marquée, celle-là – de l'emploi d'exemples attestés présentés sans contexte provient de Nagy (2017). Ici, l'auteure cite comme exemples des phrases par paquets successifs, avec des interruptions entre certains, tirées d'un roman d'Anna Gavalda. La problématique ici est de déterminer l'identité d'un personnage aux contours flous, introduit initialement (à la première phrase d'un roman) au moyen d'un pronom masculin sans antécédent, puis plus loin désigné via un pronom féminin. Il s'agirait donc du procédé *in media res*, souvent employé dans les romans contemporains :

- [6] *Il se tenait souvent à l'écart.* (Anna Gavalda, *La Consolante*, 2008, p. 9, Paris : Le [sic] *dilettante*, cité par Nagy, 2017 : 41)

Or, les emplois de *il* sans antécédent continuent jusqu'à la 38^e phrase. Nagy dénombre chaque phrase citée, mais n'indique pas à son lecteur le propos des phrases intercalées non citées. La séquence de phrases citées est celle-ci : 1-8, 56, 14, 26-48, 58-63, 66, présentée dans l'ordre, non-consécutif donc, indiqué dans le texte. Face à une telle situation, il sera fort difficile au lecteur, sinon totalement impossible, de reconstituer le *discours* associé à ce (non-)texte constitué par les citations non continues de phrases du roman de Gavalda.

À noter que, en fonction de ce qui est prédiqué du référent (encore indistinct) du pronom sujet *il* ici ('le fait de se tenir à chaque occasion à l'écart – sous-entendu, par rapport à d'autres personnes dans des réunions ou groupes collectifs'), celui-ci serait un individu

humain masculin. Ces précisions provenant du prédicateur (*se tenir à l'écart*) de la phrase donneraient lieu, à leur tour, à l'inférence selon laquelle l'individu en question serait plutôt solitaire, peu social. De plus, la représentation mentale, au départ assez floue, s'enrichira considérablement au fur et à mesure de l'arrivée du contenu des phrases ultérieures, jusqu'à celle où l'identité complète de ce référent sera assurée, mais ce sera certainement une confirmation, en quelque sorte, de ce que le lecteur aura déjà inféré en amont, comme on le verra.

Par ailleurs, comme le mystère est déjà savamment entretenu par la romancière dans le texte d'origine, le fait que Nagy ne cite que des séquences de phrases choisies çà et là dans la suite du roman, sans faire le point de l'évolution de l'intrigue dans les parties non citées, ne fait qu'ajouter à l'incertitude ressentie quant à l'identité du personnage (ainsi qu'au développement de l'histoire elle-même), et ce jusqu'à la phrase 38. Ici, soudain, le pronom féminin *elle* est employé pour renvoyer au même individu. Ce changement de genre référentiel est (apparemment) motivé via un dialogue entre deux des personnages, dont l'un n'est pas certain de savoir si l'individu aux contours flous est un homme ou une femme (phrase 32). Dans la phrase 36, l'incertitude est levée par l'annonce que la personne est « Un monsieur bien sûr. Mais qu'il appelait sa nounou » (*il* ici renvoie à 'Alexis', un personnage introduit plus tôt dans la narration). Ici, ce sont donc à la fois l'auteure citée et l'auteure citante qui s'abstiennent de fournir un contexte pour leurs textes respectifs.

Concernant la conception de l'anaphore chez Nagy, il est piquant de constater que, alors que le sous-titre du chapitre dont les citations en exemples sont tirées (le chapitre 4) est « Le pronom personnel *il* **comme élément de cohésion textuelle** dans le roman *La Consolante* d'Anna Gavalda », les emplois du pronom *il* dans les extraits cités sont très insolites, marqués, et ne sont pas vraiment 'cohésifs' du tout ! Cependant, ceci n'empêche aucunement l'auteur de se servir du dispositif de la co-indiciation pour marquer ces occurrences¹⁷.

Comme dernière illustration du choix et du traitement de données attestées, l'exemple [7] ci-dessous permet de voir comment un linguiste, même en présentant une donnée authentique (en l'occurrence, orale), peut se tromper en l'analysant purement en tant que structure abstraite plutôt que comme donnée interactive fruit d'un contexte énonciatif particulier. Nous revenons pour cela à l'ouvrage de Givón (2017), déjà rencontré aux §§ 2 et 3 *supra*.

Ainsi, au chapitre 5 de ce travail, l'auteur tente de remettre en cause la prétendue distinction au sein de la Grammaire Générative entre langues « configurationnelles » et « non-configurationnelles ». Pour ce faire, il cherche à montrer que les cases vides à la fois en position sujet et non-sujet sont fréquentes en anglais parlé – ce qui est contraire à ce qui est argumenté en GG, l'anglais y étant classifié comme langue « configurationnelle » prototypique (donc ayant un SV, permettant ainsi au linguiste de représenter sa structure syntagmatique au moyen d'une arborescence). Pour mettre à l'épreuve cette hypothèse, Givón avait fait visionner à des paires de sujets une courte vidéo (sans doute muette, sans dialogue). Il avait été indiqué au préalable à chaque sujet que la vidéo que son 'binôme' allait regarder était légèrement différente de la sienne. Chaque membre de ces paires devait donc essayer de mettre au clair, par la discussion, les similarités ainsi que les différences entre les deux films. Cependant, lors de l'analyse des données fournies par cette expérimentation, aucun compte ne semble avoir été tenu de la distinction entre les véritables occurrences de formes zéro, et les simples hésitations suivies de réparations. Un exemple provient de la section 3.4.1.5.4 du chapitre 5, portant sur les « objets d'anaphore-zéro » :

¹⁷ L'emploi que fait Nagy des indices de coréférence est hautement idiosyncratique à travers l'ensemble de l'ouvrage.

[7] *The guy in my story picked [Ø]...got [Ø]...and picked up all those tools* [CAT]¹⁸
 (Givón, 2017, ex. (34a), p. 150) ‘Le type dans mon histoire a choisi/cueilli [...] a eu [...] et a ramassé tous ces outils [...]’

Toutefois, cet extrait ne correspond aucunement à une structure propositionnelle complète telle quelle, comme le suggère l’analyse fournie : nous avons plutôt affaire ici, vraisemblablement, à une série d’hésitations de la part du locuteur concernant le verbe optimal à choisir, hésitations à l’issue desquelles il ou elle décide finalement de choisir la même forme de verbe que celui employé au départ, mais suivie d’une particule (« pick » tout seul signifie en anglais soit ‘choisir’, soit ‘cueillir’, et « pick up », ‘ramasser’). Contrairement donc à ce qui est indiqué à ce point de l’ouvrage, il n’y a pas ici de « cataphore », dispositif référentiel hautement marqué, comme le note Givón lui-même ailleurs dans l’ouvrage (2017 : 183). Le locuteur garderait donc en mémoire à court terme, pour la durée de l’énoncé, le SN objet plein *all those tools*, et aurait besoin d’un certain temps de traitement afin de rechercher le verbe optimal à employer pour le régir. À noter que le fait de remplacer les supposées « cases vides » ici par un pronom inaccentué manifeste (clitique en français) rendrait la structure agrammaticale :

[7] a. *The guy in my story picked *them_i* [...] got *them_i* [...] and picked up all those tools;
 ‘Le type dans mon histoire les_i a choisis/cueillis [...] les_i a eus [...] et a ramassé tous ces outils;_i’

Il s’agit donc certainement d’un cas de sur-analyse de la part de l’auteur, l’analyste voyant dans les productions textuelles d’un usager plus que l’on pourra raisonnablement supposer y être présent, en fonction de sa propre perspective.

En résumé, les données naturelles sont (en principe) susceptibles de révéler des faits de langue et de discours que les exemples inventés ne soulèvent guère. Mais les linguistes qui les choisissent ne fournissent pas toujours le contexte dans lequel ces extraits s’imbriquent, contexte qui est pourtant indispensable pour une saisie globale et exacte de ces données (voir les exemples [4] et [6] ci-dessus) ; et ils peuvent même se tromper dans leur analyse (le cas de l’exemple [7]), préoccupés qu’ils peuvent l’être surtout par leur volonté de démontrer telle conclusion théorique. Comme nous l’avons déjà proposé (à la fin du § 3), les données naturelles que les chercheurs recueillent pour des études de certaines procédures indexicales (anaphore, « anadeixis » et deixis) permettent en principe de postuler des théories du fonctionnement de la référence en discours bien différentes de celles classiquement proposées en fonction des données construites, écrites. Ces exemples attestés – surtout ceux produits dans des contextes d’emploi spontané, où le locuteur ou le scripteur ne prend pas garde à ce que ses énoncés soient conformes à la norme de correction langagière en vigueur – sont à même de révéler plus fidèlement la réalité de la langue en tant que système écologique¹⁹.

Cependant, de tels exemples sont parfois rejetés ou stigmatisés par certains linguistes, comme étant « marginaux » ou défectueux à certains égards (et partant, comme ne méritant pas une considération sérieuse par le linguiste). Des exemples de ce type que nous avons nous-même utilisés dans plusieurs présentations orales à des réunions de séminaires ou à des colloques ont été traités de la sorte – mais pas universellement, toutefois. Voir à cet égard les données attestées présentées et analysées dans le numéro 20 de la revue suisse *TRANEL*

¹⁸ Les symboles « Ø » sont de l’auteur. Ils sont censés marquer les « positions zéro » dans cet extrait. Par ailleurs, « CAT » signifie qu’il s’agit ici, selon l’auteur, d’un cas de « cataphore ».

¹⁹ Voir Gadet et Guerin (2016 : 4) pour l’emploi de ce terme en relation avec la constitution d’un corpus de français oral non standard chez des jeunes multiculturels à Paris. Il s’agit ici d’enregistrements de sujets sans la co-présence de l’enquêteur.

(1993) consacré aux données non standard en français (Reichler-Béguelin, éd. 1993), ainsi que dans Gadet et Guerin (2016) et Cornish (1997). Voir aussi Jefferson (2017) pour de nombreux exemples attestés d'erreurs, disfluences et ratages de communication dans des extraits d'anglais américain parlé non planifié et non standard, puis les manières dont les locuteurs ou leurs interlocuteurs les réparent. Les corpus abondent, donc, en instances où les locuteurs ou scripteurs se trompent – une illustration en serait l'extrait présenté en [9a] (§ 4.1 *infra*) –, se corrigent, ou sont corrigés par leur interlocuteur (dans le cas d'un échange conversationnel), cf. aussi Sprouse et Schütze, sous presse, p. 2. Les symboles employés pour caractériser les différents statuts des exemples inventés (voir le § 5) seraient donc valables aussi pour marquer ceux des exemples attestés.

4.1 *Étude de cas : la problématique de l'anaphore pronominale indirecte. Contextualisation et expérimentation*

Dans ce même ordre d'idées, plusieurs linguistes ont imposé des restrictions sévères sur l'emploi de pronoms inaccentués à référence anaphorique indirecte. À la différence de l'anaphore directe, où (selon la norme) l'anaphorique renvoie à une entité préalablement introduite explicitement par un antécédent textuel canonique (voir l'exemple [2b] ci-dessus), l'anaphore indirecte exige de l'interprète qu'il effectue des inférences à partir du co-texte, faisant souvent intervenir pour cela ses connaissances du monde, du genre de texte à l'œuvre, etc. Plus précisément, l'interprétation d'une anaphore indirecte exige chez l'interprète une inférence semi-automatique de « pontage » afin de passer de ce qui est explicitement mentionné ou focalisé à un référent qui y est associé d'une façon ou d'une autre – via l'invocation d'une relation de partie-à-tout, d'occurrence à type, ou métonymique. Pour que la récupération soit viable, il doit y avoir une relation topicale ou d'à-propos potentielle entre le référent indirect et le discours ultérieur. Trois exemples représentant chacun trois langues différentes (respectivement, le français, l'anglais et l'allemand) sont présentés sous [8] :

- [8] a. *Ah dis donc maman tu t'souviens Cinéma Paradiso, ben il a fait un nouveau film* (exemple (65) cité dans Reichler-Béguelin, 1993)
- b. Femme : *Why didn't you write to me?*
 'Pourquoi tu ne m'as pas écrit ?'
 Homme : *I did..., started to, but I always tore 'em up.*
 'Si, je l'ai fait..., ai commencé à le faire, mais à chaque fois je les avais déchirées' (Extrait du film *Summer Holiday*)
- c. *Ich wäre wunschlos glücklich, wenn Sie nicht immer auf dem Lehrerparkplatz parken würden. Das nächste mal lasse ich ihn abschleppen*
 'Je serais vraiment heureux/-se si vous ne vous gariez pas toujours dans le parking des professeurs. La prochaine fois, je la ferai enlever' (Exemple extrait du feuilleton de télévision *Beverly Hills 90210*, RTL allemande, 31/08/95, cité dans Consten, 2001)

Les extraits en [8a, b] représentent l'usage familial oral, mais celui en [8c] (une note écrite attachée sur le pare-brise de la voiture incriminée) relève d'un genre plus formel, soigné. Dans [8a], le référent visé est récupéré grâce à l'invocation d'une relation métonymique ('film'-'réalisateur'), dans [8b], le moyen utilisé serait la connaissance linguistique de la structure argumentale du prédicateur 'write' (au sens de 'entrer en correspondance avec quelqu'un'), puis en convoquant les connaissances du monde y afférant ; et en [8c], c'est la

combinaison de la structure argumentale du prédicateur ‘parken’ (‘se garer’) et de la perception de la voiture dont il est question dans la situation d’énonciation, qui motiverait l’inférence requise.

Or, selon Dik (1978 : 20), par exemple, « Anaphoric reference to implicit referents is the exception rather than the rule ». Pour leur part, Erkü et Gundel (1987 : 539) affirment carrément que « Indirect anaphora may not be realized via (unstressed) pronouns ». Plus tard, Gundel *et al.* (2000 : 93) soutiennent que « les anaphoriques pronominaux et démonstratifs indirects sont relativement peu fréquents par rapport aux anaphoriques indirects introduits par un article défini ou indéfini, et qu’ils apparaissent plus fréquemment dans des textes non-planifiés que dans des textes planifiés ». Dans leur conclusion, les auteurs prétendent que « les anaphoriques indirects ne peuvent, de façon typique, être codés au moyen de syntagmes pronominaux ou démonstratifs [...] »²⁰ (Gundel *et al.*, 2000 : 100).

Or, ce dernier article présente comme données à analyser, justement, des extraits attestés relevant de types d’emplois d’anglais américain non planifiés, oraux comme écrits (comme exemples, voir [9a-c] ci-dessous). Et cependant, la conclusion que les auteurs en tirent est que les exemples retenus d’emplois de pronoms ou de démonstratifs pour réaliser des anaphores indirectes sont marginaux (et donc théoriquement inintéressants), et ne sont donc pas de nature à mettre en cause la célèbre « Givenness Hierarchy » (Gundel *et al.*, 1993 : 275 ; reproduite dans Gundel *et al.*, 2000 : 82, item (8)). Celle-ci place les pronoms inaccentués en tête de l’échelle, où ils sont censés encoder uniquement le statut cognitif « en focus », et non celui d’« activé » (la position immédiatement à droite de celle-ci). Le statut « en focus » correspond à l’activation la plus élevée de la représentation mentale d’un référent : c’est celle à laquelle les participants prêtent une attention mutuelle centrale au moment de la parole. Le statut d’« activé », lui, représente un degré d’activation attentionnelle moins élevé : la représentation concernée se trouve à la périphérie de leur conscience au moment de l’énonciation.

En réalité, le propre des anaphores indirectes est bien de cibler un référent qui est cognitivement « activé » (comme c’est le cas des trois exemples en [8a-c] ci-dessus), mais pas nécessairement « en focus » – leur interprétation exigeant une inférence quasi automatique à chaque fois.

Quant aux exemples présentés dans Gundel *et al.* (2000), ils peuvent être analysés pour l’essentiel comme étant produits, soit parce que, pour le locuteur ou le scripteur, leur référent est hautement saillant mais qu’il ne se met pas à la place de son interlocuteur ou lecteur afin de mieux motiver son renvoi dans son contexte ; soit parce qu’il s’agit d’un emploi général (le *they* dit « collectif » – référence non spécifique, donc -, par exemple) où l’identité du référent n’est pas jugée importante pour la saisie du propos ; soit encore parce que l’émetteur considère que dans le contexte à l’œuvre, son allocataire ou lecteur pourra aisément récupérer le référent visé en fonction d’une expression fort atténuée, en se fondant sur la loi du moindre effort. [9a-c] illustrent respectivement ces trois cas de figure :

- [9] a. *My mum bought an exercise tape and so I’ll go nuts and play it in the morn and in the afternoon and do the added things she says for those that want a more intense workout.*

²⁰ En fait, les occurrences de marqueurs indexicaux (pronoms de 3^e personne, pronoms démonstratifs, voire même également, formes nulles) sans antécédent textuel sont bien plus fréquentes dans l’usage que les auteurs mentionnés à l’instant ne l’affirment (dans certains genres et sous-genres plus que dans d’autres – notamment, dans la conversation spontanée (cf. l’anaphore indirecte), les brèves de faits divers des journaux, les résumés de films, la publicité, les titres de journaux, et même dans des romans ou autobiographies). Voir à ce sujet Gerrig *et al.* (2011) et Kitzinger *et al.* (2012), le dernier surtout pour une analyse des stratégies communicatives que ce phénomène peut permettre de réaliser. Voir Cornish (1996, 2007) pour un passage en revue plus général.

‘(Ma) maman a acheté une cassette de gym et donc je deviens folle, et la passe le matin comme l’après-midi et fais toutes les choses supplémentaires qu’*elle* spécifie pour ceux qui veulent des séances d’entraînement plus intensives’ (soutien. alt. désordres d’alimentation ; ex. (26) dans Gundel *et al.*, 2000 : 96) [*she* = ‘l’institutrice sur la cassette de gym’]

- b. *It is very hard for me to feel supported after recently being discharged from an intensive treatment program. Today I got weighed and I gained a quarter of a pound and they think that I water loaded!! ha!*

‘Il m’est très difficile de me sentir soutenue après avoir été récemment renvoyée d’un programme de traitement intensif. Aujourd’hui je me suis fait peser et j’ai pris 113 grammes et *ils* pensent que je me suis gonflée d’eau !! heuh !’ (soutien. alt. désordres d’alimentation ; ex. (23) dans Gundel *et al.*, 2000 : 95)

- c. *Seven years of marriage. Yes we had our ups and downs but now she says she doesn’t love me anymore.*

‘Sept ans de mariage. Oui nous avons eu des hauts et des bas, mais maintenant *elle* dit qu’*elle* ne m’aime plus’ (soutien alt. divorce ; ex. (28) dans Gundel *et al.*, 2000 : 97)

Gundel *et al.* (2000) sont en fait d’accord avec Sanford *et al.* (1983) pour qualifier les genres oraux et écrits non planifiés de discours « relâchés », où des libertés sont prises par rapport à la norme de correction manifestée par l’écrit « explicite ». C’est la raison pour laquelle les auteurs optent pour une solution selon laquelle les extraits de ce type qu’ils présentent ici ne constituent que des « violations mineures » de la Givenness Hierarchy (2000 : 100), qui à ce titre ne méritent pas une mise en cause de ses prédictions. Or, on a là un exemple flagrant de ce que Sampson (2017 : 4) qualifie de « scientisme » :

« The trouble with applying this idea [la différence selon K. Popper entre les sciences et les non-sciences] to the study of language is that human language behaviour is a very open-ended activity. As with some other aspects of human life, as fast as one tries to draw boundaries round sets of “things that can happen”, to contrast them with other imaginable “things that can’t happen”, we find that human behaviour breaks through the boundaries ».

Cette attitude correspond très exactement à la première des deux positions possibles, selon Willems (1985 : 94), adoptées par des théories « qui se veu[le]nt explicative[s] et explicite[s] » : à savoir, le rejet des faits d’exception « en dehors du domaine étudié ».

De façon analogue, mais dans le domaine de la psycholinguistique à présent, Sanford et Garrod (1981) soutiennent, tout comme Erkü et Gundel (1987), que les pronoms « sans antécédent » seraient carrément impossibles. Mais les auteurs ne distinguent pas entre des référents « activés » ou relativement centraux psychologiquement dans tel ou tel contexte, et ceux qui ne sont que « périphériques » (relevant du moyen ou de l’instrument par lequel la situation évoquée aura été mise en place : un exemple en serait l’interprétation du pronom *it* dans [10], phrase 2, dans la version « antécédent implicite »). Leurs exemples de « pronoms sans antécédent illicites » relèvent toujours de ce dernier type, et non du premier – caractérisé, lui, par les pronoms figurant dans les exemples [8a-c] et [9c] ci-dessus. Le statut cognitif des référents périphériques correspondrait à celui de ‘familier’ (un cran plus bas, à droite, donc) dans l’Échelle du donné, selon Gundel *et al.* (1993).

Sanford *et al.* (1983) ont voulu mettre cette thèse à l’épreuve. Ils avaient donc commencé par administrer à 30 sujets une tâche préliminaire de lecture auto-segmentée en

différé de plusieurs séries de petits textes écrits comprenant chacun trois phrases (la première établissant une situation mise en place grâce à un instrument quelconque (soit mentionné explicitement, soit laissé implicite), la seconde comportant une reprise de cet instrument via un pronom, et la troisième représentant un commentaire de la situation évoquée précédemment). L'exemple [10], tiré des matériaux utilisés par Sanford *et al.*, 1983, Experiment 1, en est une illustration :

[10]	<u>Explicit antecedent</u>	<u>Implicit Antecedent</u>
Sentence 1	Roland parted his long hair with a comb. 'Roland s'est fait une raie dans ses longs cheveux avec un peigne'	Ronald parted his long hair. 'Roland s'est fait une raie dans ses longs cheveux'
Sentence 2 (Target)	It was twisted with many teeth missing. 'Il était tordu et beaucoup de dents manquaient'	
Sentence 3	He had had it since childhood. 'Il l'avait depuis son enfance'	
Question	Did Ronald part his hair with a brush? (No) 'Ronald s'est-il fait une raie avec une brosse ? (Non)'	

Dans cette expérimentation en temps réel, seule une différence (non-significative) de 139 ms fut obtenue entre les temps de lecture moyens pour les phrases pronominales cibles (la seconde à chaque fois) dans la condition « antécédent explicite » et dans celle de l'antécédent implicite (pour les cas où aucun autre « antécédent possible » de même genre et/ou nombre n'était disponible dans la phrase de départ²¹).

Or, suite à cette expérimentation en ligne, les auteurs avaient administré aux sujets testés un ensemble de textes écrits, réunis dans un livret. Ces textes étaient composés de chacune des conditions représentées dans l'expérimentation 1, les mêmes matériaux étant employés. L'instruction proposée aux sujets sous-entendait une attitude quelque peu normative vis-à-vis de ces textes : « Ré-écrivez ces textes si vous le jugez nécessaire, **de façon à les rendre 'plus acceptables'** » (« [...]Ainsi, comme on pouvait le prédire, seuls 7% des textes comportant un antécédent explicite avaient été ré-écrits, alors qu'entre 83% et 92% de ceux ayant un « antécédent » implicite l'avaient été. Ces résultats peuvent être compris comme reflétant l'influence de ce contexte en différé, réflexif, ainsi que de l'effet implicitement normatif de l'instruction (où l'implication créée est que 'ces textes peuvent ne pas tous être tout à fait acceptables'). Il y a d'ailleurs une contradiction flagrante en termes de genre au sein de ce texte lui-même : car il s'agit ici d'une description (typique de l'écrit normé) d'une situation. Mais dans la version « implicite », l'emploi du pronom *it*, ostensiblement en anaphore indirecte, sous-entendrait soudain un genre non planifié, typique de l'oral familier.

Il s'avère donc que la conclusion tirée des résultats de ce test et de cette expérimentation par les auteurs n'est guère valable. À savoir que « This study clearly demonstrates an unprompted judgement that pronouns without antecedent are unacceptable – at least in materials of the type used here » (Sanford *et al.*, 1983 : 313). On a affaire ici en réalité à des genres distincts relevant de l'oral spontané, familier en situation (les exemples « motivés » [8a, b] et [9c]), d'une part, et de la prose écrite soutenue, de l'autre, ainsi qu'à des

²¹ Cela est bien le cas en [10], où le pronom *it* dans la phrase critique cible 2 pouvait être apparié ("bonded"), de façon inappropriée, au référent 'Roland's long hair' ('les longs cheveux de Roland'), puisque le SN *his long hair* dans la phrase 1 de ce texte est à la fois au singulier et neutre (inanimité). En anglais, à la différence du français *cheveu*, le nom *hair* est massif, non comptable. Cependant, lorsqu'on prend en compte (comme cela sera nécessaire, de toute façon) la nature de ce qui est prédiqué du référent potentiel du pronom *it* ici, un tel candidat serait exclu : des 'cheveux' peuvent bien être *twisted* ('tordus'), mais ne contiennent certainement pas des *teeth* ('dents'). Ces deux propriétés conviendront parfaitement à un 'peigne' (*comb*), par contre. Un appariement de *it* avec le référent 'Roland's long hair' aurait donc certainement donné lieu à un « garden path », une « impasse » interprétative.

attentes du lecteur ou de l'auditeur fondées sur leur expérience des contextes convenant à ces genres. Voir à ce sujet Oakhill et Garnham (1992).

Quoi qu'il en soit, les expérimentations en anglais et en français rapportées dans Cornish *et al.* (2005) indiquent clairement que, lorsqu'on distingue entre référents implicites « centraux, nucléaires » et « périphériques », les pronoms reprenant ceux du premier type (à la différence du second) sont compris presque aussi rapidement dans la condition « antécédent implicite » que dans l'« explicite » et ce dans les deux langues testées. Ces résultats ont été confirmés plus récemment par une autre série d'expérimentations avec l'allemand comme langue de base (cf. Järvikivi *et al.*, 2019). Autre différence par rapport aux expérimentations de Sanford et ses associés : les matériaux utilisés étaient présentés sous forme de dialogues, ce qui rendait sans doute plus naturelle la présence de pronoms indirects. Voir aussi les nombreux autres exemples attestés de ce phénomène analysés dans Conte (1991) pour l'anglais, l'allemand et l'italien, Reichler-Béguelin (1993) pour le français, ainsi que McCawley (1991) pour l'anglais.

En somme, les exemples qui peuvent paraître « extrêmes » à certains égards (tout comme les exemples franchement inacceptables) sont utiles dans la mesure où ils nous renseignent sur la présence et sur l'absence de telle ou telle propriété intrinsèque d'une structure ou d'un type d'énoncé, ce qu'une concentration uniquement sur des données « standard » ou « canoniques » ne peut guère offrir.

5. Le statut des exemples présentés pour l'analyse : grammaticalité, sémantique, cohésion textuelle et cohérence discursive

Le statut que le linguiste attribue aux exemples proposés pour l'analyse est étroitement lié au choix et à la nature de ceux-ci. Comme nous l'avons vu au § 2 *supra*, ce statut est manifesté dans la plupart des cas par un symbole conventionnel préfixé directement aux données présentées elles-mêmes : l'astérisque (parfois questionné) pour signaler l'agrammaticalité, le point d'exclamation pour une anomalie d'ordre sémantique dans l'interprétation d'un exemple, et le dièse pour marquer que son emploi en tant qu'énoncé (potentiel) dans tel ou tel contexte d'énonciation est (ou serait) peu naturel. Un autre symbole employé, lui, dans le cas d'énoncés attestés dans les corpus est le symbole de pourcentage ('%'). Celui-ci sert à indiquer que le type d'exemple ainsi préfixé se trouve chez nombre de locuteurs dans tel corpus d'énoncés, mais non chez tous. Et nous avons déjà mentionné (au § 3) le symbole '~~' proposé par Chafe (1994) pour caractériser certains exemples inventés dans la littérature qu'il serait peu probable de relever dans les corpus d'usages.

Cependant, dans la pratique, les trois premiers statuts ne sont pas toujours clairement distingués. Dans le cas des travaux d'ordre formaliste (syntaxe et sémantique formelles), par exemple, il est fréquent que le linguiste adopte très généralement l'astérisque pour marquer toute sorte de déféctuosité remarquée dans un exemple, qu'elle soit d'ordre syntaxique, sémantique ou discursif. Toutefois, à ce titre, une telle généralisation revient à neutraliser complètement la valeur et la fonction des exemples étoilés par ceux qui les emploient à ce titre. En voici quelques illustrations :

[11] *Reagan_i is aware that the former actor_i is regarded by many with a good deal of skepticism.

'Reagan_i est bien conscient que l'ancien acteur_i est considéré par beaucoup avec une grande dose de scepticisme' (Chomsky, 1986 : ex. (49), p. 80)

Il aurait été préférable en fait de préfixer le dièse (modifié par le diacritique ^{CRD} : voir le § 2) plutôt que l'astérisque qu'emploie Chomsky pour cet exemple afin de caractériser son statut.

Après tout, il n'y a rien de problématique en [11] en tant que *phrase*. C'est plutôt au niveau du *discours* auquel il pourra donner lieu par défaut (en tant qu'énoncé potentiel) qu'il pose problème. Car ici, ce qui provoque le caractère inacceptable de l'exemple est le fait que le contenu de la subordonnée est présenté du point de vue de Reagan lui-même, en tant que sujet de conscience de l'ensemble de ce fragment d'un discours potentiel ('Reagan is aware that...'). L'emploi du pronom *he* à la place de *the former actor*²² aurait été parfaitement cohérent (voir aussi à ce propos l'exemple inventé [2b] par contraste avec [2a] au § 3). Telle quelle, donc, on est en présence en quelque sorte d'une situation de « fils perspectivaux croisés ou emmêlés ». De ce fait, on pourrait la qualifier de « dissonance polyphonique » (tout comme l'exemple [2a]). Le statut d'agrammaticalité marqué par l'astérisque que Chomsky préfixe à cet exemple découle du fait que, selon le modèle « Gouvernement et Liage » en vigueur à l'époque, il contrevient au Principe C du « Liage », qui prescrit que « les expressions référentielles » [dont fait partie *the former actor*] sont libres [donc non liées] partout ».

- [12] **He was hit on the head before the lecturer had a chance to say anything.*
(Lit.) 'Il fut frappé à la tête avant que le conférencier n'ait pu dire quoi que ce soit' (Reinhart, 1983 : ex. (28a), ch. 2, p. 37)

Cet exemple inventé provient d'une étude sur l'interprétation des différents sous-types d'anaphore pronominale en anglais, dans le cadre de la grammaire générative. Tanya Reinhart, l'auteure, cependant, y accorde une assez grande place à des considérations discursives, sémantiques et pragmatiques, plus encore que dans aucune autre étude précédente faite par un(e) Générativiste. Reinhart emploie déjà le dièse en même temps que l'astérisque pour marquer les statuts de certains de ses exemples (tous inventés) ; mais elle ne distingue pas toujours à bon escient entre l'attribution de ces deux statuts.

L'exemple [12] en est d'ailleurs une illustration. La présence d'un astérisque devant cet exemple peut être imputée à son analyse selon laquelle le pronom sujet *he* ici c-commande son antécédent putatif *the lecturer*²³. Cependant, il semble clair qu'en réalité cet exemple est bien formé. Car la principale anaphorique en [12] peut être analysée comme étant pragmatiquement subordonnée par rapport à celle contenant l'« antécédent » (voir Mittwoch, 1983 pour d'autres exemples de ce cas de figure). La présence devant [12] d'un astérisque est donc injustifiée. Il y a un certain nombre d'autres exemples présentés comme « agrammaticaux » dans cet ouvrage, exemples qui s'avèrent en réalité parfaitement recevables.

Voici un troisième exemple, extrait comme [12] de Reinhart (1983) :

- [13] **In Ben's office, he placed his new brass bed.*
'Dans le bureau de Ben, il²⁴ a installé son nouveau lit en cuivre'

²² Ce SN défini représente une description « objective » (donc « externe ») du référent 'Ronald Reagan', et ne serait certainement pas (dans ce contexte, du moins) un qualificatif qu'il aurait employé lui-même pour se référer (emploi « logophorique », donc). C'est à l'évidence le locuteur / scripteur de ce fragment, en tant qu'énoncé, qui l'aura pris en charge.

²³ Dans l'approche classique de la théorie du Gouvernement et du Liage, si un pronom c-commande son antécédent potentiel au sein d'une configuration syntaxique quelconque, dans ce cas il ne peut pas être « lié » par lui (autrement dit, aucune relation anaphorique ne pourra être établie entre les deux expressions, la c-commande étant une relation entre nœuds à l'intérieur d'une structure phrasique donnée). En effet, dans [12], le premier nœud branchant dominant ce pronom domine également *the lecturer*, dans un arbre représentant la structure syntaxique de l'exemple : car la proposition dont *he* est le sujet est ici la principale, et celle dont *the lecturer* est sujet lui est subordonnée.

²⁴ À noter ici qu'il aurait sans doute été plus naturel en français d'employer un pronom démonstratif à la place du pronom clitique de 3^e personne *il* : à savoir, *celui-ci*.

(Reinhart 1983 : ex. (31b), ch. 3, p. 68)

Ici aussi, le pronom *he* c-commande l'« antécédent » *Ben*, même si dans ce cas-ci il le suit au plan linéaire. Pourtant, une autre explication qu'une simple violation d'une contrainte syntaxique est préférable ici. À savoir que nous sommes en présence d'une situation d'incohérence référentielle (au niveau énonciatif, donc). Car le SP antéposé *in Ben's office* constitue à ce niveau un cadre référentiel. À ce titre, ce n'est pas à l'individu nommé « Ben » en chair et en os (en tant qu'agent potentiel, donc) qu'on renvoie, mais à une caractéristique du bureau dont il est question, afin de permettre au destinataire de le situer (c'est celui que le dénommé « Ben » occupe ou qui lui appartient). Cependant, dans la principale qui suit ce cadratif, il est question d'une action dont cet individu serait l'agent. Il s'agit donc d'une confusion de rôles et de plans, en quelque sorte : le référent du pronom *he* est à la fois censé être l'agent de la mise en place d'une action au premier plan de ce discours, et tirer sa référence en fonction d'un aspect de la localisation d'une entité mise en arrière-plan de ce même discours (voir Cornish, 2012 pour une discussion plus ample de ce type de configuration). Comme [11], l'exemple mérite donc plutôt le dièse avec la mention ^{CRD} que l'astérisque.

Voici un dernier exemple inventé :

- [14] John and Mary came into the room. He was laughing but she was crying.
(Lit.) 'John et Mary sont entrés dans la pièce. Il riait mais elle pleurait' (Ariel, 1990 : 52, ex. (6))

Ici, à la différence de [11], [12] et [13], il n'y a pas de symbole préfixant l'exemple du tout. Cependant, pour être recevable en tant que fragment de discours possible, une intonation contrastive doit être surimposée aux pronoms *he* et *she* dans la seconde phrase – les éléments prédicatifs des deux propositions dans celle-ci étant distincts, attribuant donc une propriété différente à chaque individu évoqué via les pronoms sujets, le contraste entre les deux propositions de la seconde phrase étant également marqué via le connecteur *but* qui les coordonne. Sinon, tel quel (avec les pronoms sujets par défaut inaccentués), cet exemple aurait dû être marqué par un dièse, cette fois avec l'indication ^{CST} 25. Il s'agit donc ici d'un problème relevant du *discours* et non de la *phrase* en tant que telle, l'intonation étant un moyen de contextualisation par excellence.

La pratique consistant à marquer les exemples inacceptables de types variés par un astérisque sous-entend que pour l'auteur, c'est la syntaxe qui prime sur les autres dimensions d'un énoncé, ce qui est loin d'être la position adoptée par la grande majorité des autres cadres linguistiques. C'est précisément la linguistique générative chomskienne qui représente le plus centralement cette notion, les composantes phonologiques et sémantiques dans une grammaire générative étant classiquement « interprétatives » du « squelette » formel constitué par la représentation syntaxique d'une phrase à générer.

Mais parfois, on tombe sur cette même pratique chez des auteurs se situant bien loin de cette mouvance : une illustration se trouve chez le cognitiviste Talmy (2013 : 99), qui soutient que « if there is a scattering of books in front of me and I just look directly at you and say *That's my book, that is an ill-formed sentence. It is ungrammatical* » (c'est moi qui

²⁵ Car ici on est en présence d'une déviance de type textuel, impliquant donc la cohésion plutôt que la cohérence. Par convention, l'intonation contrastive sur un élément segmental est marquée par les majuscules. Donc *HE* et *SHE* dans [14]. Dans la version française de cet exemple, on aurait une paire de pronoms disjoints, et non clitiques : à savoir, *LUI* et *ELLE*, respectivement (je suis redevable à Anne-Catherine Simon (c.p) pour ce rappel ; selon celle-ci, cependant, la version comprenant deux pronoms clitiques en français serait recevable, mais moins naturelle toutefois que celle avec les pronoms disjoints accentués).

souligne). Car pour Talmy, le geste attendu indiquant le livre visé est absent. Pourtant, ceci relève bien entendu de la situation d'énonciation à l'œuvre ici, et non de la « phrase » *That's my book* en tant que telle.

Néanmoins, il faut reconnaître qu'il y a eu une évolution dans les études formelles depuis le début des années 2000 : l'emploi du dièse s'est beaucoup généralisé dans les travaux prenant en compte des suites de phrases, même inventées. C'est le cas par exemple de l'article de Roussarie et Amsili (2007), qui porte sur l'interprétation des « pronoms phrastiques » en français.

6. Vers une conclusion

Au terme de cette investigation, tentons de dresser un bilan.

Tout d'abord, en ré-analysant un certain nombre des exemples proposés par les linguistes (tant inventés qu'attestés) dans le but de décrire et/ou d'expliquer un phénomène textuel ou discursif, il est apparu une certaine confusion entre des facteurs qu'il appartient de maintenir séparés. Les confondre (ou bien négliger certains d'entre eux) brouille les faits langagiers que l'auteur-linguiste cherche à établir, et par conséquent affaiblit d'autant, selon nous, la validité de ses conclusions.

D'abord, certains exemples inventés uniquement pour les besoins de justification d'une théorie manquent de naturel, ce qui rend difficile la recherche d'un contexte qui pourrait les motiver, permettant ainsi aux sujets-locuteurs de la langue en question de les évaluer. C'est notamment le cas des deux exemples anglais en [1a,b], puis de l'exemple français en [2a] (au § 3). Nous avons vu que, pour les deux premiers, soit l'une des constructions syntaxiques qu'ils manifestent ne convenait pas au co-texte à l'œuvre, soit la déviance se situait au niveau de la structure informationnelle. Pour le second, au contraire, la gêne ressentie pouvait être mise sur le compte de l'absence de motivation de la reprise d'un référent fraîchement introduit au sein du co-texte, par un marqueur référentiellement autonome à l'endroit où celui-ci apparaît dans ce fragment de texte ; puis (et surtout) sur celui de l'apparente confusion polyphonique concernant la source du point de vue caractérisant chaque segment, confusion liée directement au choix de ce marqueur.

Ces trois exemples n'étaient pourtant précédés d'aucun symbole indiquant un aspect problématique de sa construction syntaxique ou sémantique interne, ni de son emploi potentiel dans un contexte d'usage particulier²⁶. Par contre, en [11], [12] et [13] (§ 5), ces exemples (tous également inventés) sont bien précédés d'un tel symbole, à savoir l'astérisque. Cependant, en y regardant d'un peu plus près, il s'est avéré que le problème ne relevait pas du tout de la syntaxe en tant que telle, mais plutôt de facteurs ayant tous trait à l'énonciation : un croisement de fils perspectivaux (un cas de « dissonance polyphonique ») dans [11], transformé en énoncé potentiel, et dans [13], malgré la présence d'un astérisque, un problème relevant de l'intégration cohérente du contenu des deux constituants majeurs de la phrase, reliés en fonction d'une relation anaphorique. En [12] par contre, l'exemple ne présentait aucune difficulté en tant qu'énoncé potentiel, en dépit de sa violation d'une contrainte syntaxique supposée.

Or, tous les problèmes relevés ci-dessus partagent la même caractéristique fondamentale : les exemples concernés sont difficiles à contextualiser en tant qu'énoncés potentiels. Et pourtant, cela peut également être le cas (mais pour des raisons différentes) pour certains exemples attestés, extraits de textes écrits ou bien oraux. Les exemples [4] et [6] (§ 4) constituent des fragments de textes écrits qui ressemblent fort à des exemples inventés : ceux-

²⁶ Mais ceci, bien entendu, ne fait pas partie habituellement du souci des linguistes employant de façon centrale, voir exclusive, les exemples inventés.

ci sont soit extraits d'un texte source plus long, mais contiennent de nombreux aspects non autonomes contextuellement (le cas de [4]), soit les phrases citées en exemples ne sont pas toutes consécutives, avec omission des phrases intercalées (le cas de [6]). Il faudra donc dans ces cas de figure que le lecteur « re-contextualise » les extraits proposés, dans la mesure du possible, en fonction du seul co-texte présenté, opération par définition problématique dans ces circonstances. C'est là très exactement ce qu'il se voit obligé de faire dans le cas des exemples inventés.

Une autre caractéristique importante dans l'évaluation d'une donnée langagière a trait à la prosodie. En effet, la première étape vers une contextualisation réussie d'un exemple, même écrit, serait la possibilité ainsi que la probabilité de la surimposition d'une structure prosodique particulière aux éléments textuels impliqués. Ceci sera fonction de la structure informationnelle de chaque composante d'un exemple multi-propositionnel, et permettra à son tour de dégager les unités discursives auxquelles celui-ci pourra correspondre. On l'a surtout vu en caractérisant le second SN relationnel défini employé dans l'exemple construit [2a] (section 3) dans une fonction de reprise, en termes de sa possible réalisation prosodique ; puis dans l'exemple également inventé [14] de la section 5. Voir aussi Sabio (2006 : 175), qui distingue deux types d'antéposition d'objets directs en français parlé en fonction de l'intonation avec laquelle ceux-ci sont prononcés.

Enfin, la restriction des données présentées à des exemples construits à caractère écrit, normé, donne lieu, comme nous l'avons vu, à une caractérisation trompeuse d'au moins un phénomène discursif, à savoir l'anaphore. Or, une prise en compte de données échappant au contrôle par la norme régissant les textes écrits permet en réalité de saisir le mécanisme en jeu d'une manière plus fine et plus juste. Et pourtant nombre de linguistes, du fait de leur éloignement de la norme en vigueur, rejettent de telles productions comme étant peu intéressantes au niveau théorique et donc peu viables pour ce qui concerne la description d'un tel phénomène.

Références bibliographiques

- ARIEL M., 1990, *Accessing Noun-Phrase Antecedents*, Londres, Routledge.
- BADER M. et HÄUSSLER J., 2010, « Toward a model of grammaticality judgements », *Journal of Linguistics*, 46(3), p. 273-330.
- BARSS A., 2003, « Timing puzzles in anaphora and interpretation », in BARSS A., *Anaphora. A Reference Guide*, Malden/Oxford/Melbourne/Berlin, Blackwell Publishing, p. 1-22.
- BLAKEMORE D., 2010, « Communication and the representation of thought : The use of audience-directed expressions in free indirect thought representations », *Journal of Linguistics*, 46(3), p. 575-599.
- BLANCHE-BENVENISTE C., 1990, *Le français parlé. Études grammaticales*, Paris, CNRS Éditions.
- CHAFE W., 1994, *Discourse, Consciousness and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*, Chicago, University of Chicago Press.
- CHAFE W. (éd.), 1980, *The Pear Stories: Cognitive, cultural and linguistic aspects of narrative production*, Norwood, NJ, Ablex.
- CHOMSKY N., 1986, *Knowledge of Language*, New York, Praeger.
- CISLARU G. et SITRI F., 2009, « Texte et discours. Corpus, co-texte et analyse automatique du point de vue de l'analyse de discours », *Corpus*, 8, p. 29-51.
- CONSTEN M., 2001, « Indirect anaphora and deixis in texts: domain-bound reference and coherence ». Présentation d'un poster lors de l'Atelier international sur la référence, Utrecht/NL, le 11 janvier 2001.

- CONTE M-E., 1991, « Anaphores dans la dynamique textuelle », *Cahiers de Praxématique*, 16, p. 11-33.
- CORNISH F., 1987, « Anaphoric pronouns: under linguistic control, or signalling particular discourse representations? », *Journal of Semantics*, 5(3), p. 207-234.
- CORNISH F., 1996, « 'Antecedentless' anaphors: deixis, anaphora, or what? Some evidence from English and French », *Journal of Linguistics*, 32, p. 19-41.
- CORNISH F., 1997, « Non-standard anaphora, discourse integration, and coherence », *Verbum*, 19(1-2), p. 5-23.
- CORNISH F. 2007, « Indirect pronominal anaphora in English and French: marginal rarity, or unmarked norm? Some psycholinguistic evidence », in SCHWARZ-FRIESEL, M., CONSTEN, M. et KNEES, M., *Anaphors in Text. Cognitive, formal and applied approaches to anaphoric reference*, Amsterdam/Philadelphia, John Benjamins, p. 21-36.
- CORNISH F., 2012, « Micro-syntax, macro-syntax, foregrounding and backgrounding in discourse: When indexicals target discursively subsidiary information », *Journal of Belgian Linguistics*, 26, p. 6-34.
- CORNISH F., GARNHAM A., WIND COWLES H., FOSSARD M. et ANDRÉ V., 2005, « Indirect anaphora in English and French: A cross-linguistic study of pronoun resolution », *Journal of Memory and Language*, 52, p. 363-376.
- CORNISH F. et SALAZAR ORVIG A., 2016, « A critical look at the notion 'pro-form'. Evidence from indexical markers, spoken discourse and (French) child language », *Language Sciences*, 54, p. 58-76.
- COSME C., 2008, « A corpus-based perspective on clause linking patterns in English, French and Dutch », in FABRICIUS-HANSEN C. et RAMM W., *'Subordination' vs. 'Coordination' in Sentence and Text. A cross-linguistic perspective*, Amsterdam/Philadelphia, John Benjamins, p. 89-114.
- DIK S. C., 1978, *Stepwise Lexical Decomposition*, Lisse, Peter de Ridder Press.
- DIK S. C., 1997, *The Theory of Functional Grammar*, in HENGEVELD K., *The Theory of Functional Grammar. Part 1: The Structure of the Clause*, Berlin and New York, Mouton de Gruyter.
- ERKÜ F. et GUNDEL J. K., 1987, « The pragmatics of indirect anaphors », in VERSCHUEREN J. et BERTUCCELLI-PAPI M., *The Pragmatic Perspective: Selected Papers from the 1985 International Pragmatics Conference*, Amsterdam, Amsterdam, John Benjamins, p. 533-545.
- FOX B.A., 1987, *Discourse Structure and Anaphora*, Cambridge, Cambridge University Press.
- GADET F., 1990, « Les outils grammaticaux au risque de l'oral non standard », *Travaux de Linguistique*, 21, p. 13-24.
- GADET F. et GUERIN E., 2016, « Construire un corpus pour des façons de parler français non standard : « Multicultural Paris French » », *Corpus*, 15 (revue en ligne) [URL: <http://journals.openedition.org/corpus/3049>]
- GERRIG R.J., HORTON W.S. et STENT A., 2011, « Production and comprehension of unheralded pronouns: A corpus analysis », *Discourse Processes*, 48, p. 161-182.
- GIVÓN T., 2017, *The Story of Zero*, Amsterdam/Philadelphia, John Benjamins.
- GREIDANUS T. et CALLEBAUT B., 1990, « Le recours aux corpus oraux dans les analyses syntaxiques et pragmatiques », *Travaux de Linguistique*, 21, p. 81-92.
- GRICE H.P., 1975, « Logic and conversation », in COLE P. et MORGAN J., *Syntax and Semantics, Speech Acts*, vol. 3, New York, Academic Press, p. 41-58.
- GUNDEL J. K., HEDBERG N. et ZACHARSKI R., 1993, « Cognitive status and the form of referring expressions in discourse », *Language*, 69(2), p. 274-307.

- GUNDEL J. K., HEDBERG N. et ZACHARSKI R., 2000, « Statut cognitif et forme des anaphoriques indirects », *Verbum*, 22 (1), p. 79-102.
- HANKS P., 2013, *Lexical Analysis: Norms and exploitations*, Cambridge, Mass., MIT Press.
- HENGEVELD K. et MACKENZIE J.L., 2008, *Functional Discourse Grammar. A Typologically-based Theory of Language Structure*, Oxford, Oxford University Press.
- HUANG Y., 2000, *Anaphora. A Cross-linguistic Study*, Cambridge, Cambridge University Press.
- JÄRVIKIVI J., SCHIMKE S. et PYYKKÖNEN-KLAUCK P., 2019, « Understanding indirect reference in a visual context », *Discourse Processes* 56(2), p. 117-135.
- JEFFERSON G. (éds. DREW, P. et BERGMANN, J.), 2017, *Repairing the Broken Surface of Talk. Managing problems in speaking, hearing and understanding conversation*, Oxford, Oxford University Press.
- KITZINGER C., SHAW R. et TOERIEN M., 2012, « Referring to persons without using a full-form reference: Locally initial indexicals in action », *Research on Language and Social Interaction*, 45(2), p. 116-136.
- LALLOT J., 2007, « 'Dis-moi comment tu traites les exemples, je te dirai quel grammairien tu es'. Application à Apollonius Dyscole (*Syntaxe*) », *Langages*, 166, p. 58-70.
- LEE E., 2017, « Discourse properties of *now* », *Journal of Linguistics*, 53, p. 613-640.
- LEEMAN-BOUX D., 1994, *Les fautes du français existent-elles ?*, Paris, Seuil.
- LINELL P., 2005, *The Written Language Bias in Linguistics. Its Nature, Origins and Transformations*, Abingdon, Oxon/New York, Routledge.
- LUZZATI D., 1985, « Analyse périodique du discours », *Langue Française*, 65, p. 62-73.
- MCCAWLEY J. D., 1991, *A Linguistic Flea Circus*, Bloomington, Indiana IULC miméographe.
- MILNER J-C., 1982, *Ordres et raisons de langue*, Paris, Seuil.
- MITTWOCH A., 1983, « Backward anaphora and discourse structure », *Journal of Pragmatics*, 7 (2), p. 129-139.
- NAGY A., 2017, *Des pronoms au texte: Études de linguistique textuelle*, Debrecen University Press.
- OAKHILL J. et GARNHAM A., 1992, « Linguistic prescriptions and anaphoric reality », *Text*, 12, p. 161-182.
- REICHLER-BEGUELIN M.-J., 1993, « Anaphores associatives non-lexicales : incomplétude macrosyntaxique ? », in KAROLAK S. et MURYN T., *Complétude et incomplétude dans les langues romanes et slaves. Actes du VI^e colloque international de linguistique romane et slave*, Cracovie, 29 sept.- 3 oct. 1991, p. 327-379.
- REICHLER-BEGUELIN M.-J. (éd.), 1993, *Le traitement des données linguistiques non-standard*, Neuchâtel, Institut des Sciences du Langage et de la Communication, Université de Neuchâtel.
- REINHART T., 1983, *Anaphora and Semantic Interpretation*, London/Canberra, Croom Helm.
- ROUSSARIE L. et AMSILI P., 2007, « Interprétation des pronoms phrastiques », *Modèles Linguistiques*, 56 (28-2), p. 82-112.
- SABIO F., 2006, « L'antéposition des compléments dans le français contemporain. L'exemple des objets directs », *Lingvisticae Investigationes*, 29 (1), p. 173-182.
- SAMPSON G., 2017, *The Linguistics Delusion*, Sheffield, Equinox Publishing Ltd.
- SANFORD A. J. et GARROD S. C., 1981, *Understanding Written Language. Explorations in Comprehension beyond the Sentence*, Chichester/New York, John Wiley.
- SANFORD A. J., GARROD S. C., LUCAS A. et HENDERSON R., 1983, « Pronouns without explicit antecedents? », *Journal of Semantics*, 2, p. 303-318.
- SCHÜTZE C.T., 2016, *The Empirical Base of Linguistics. Grammaticality Judgments and Linguistic Methodology*, Berlin, Language Science Press.

- SCHÜTZE C.T. (à paraître). « Acceptability ratings cannot be taken at face value », in SCHINDLER S., DROZDOVICZ A. et BRØCKER K., *Linguistic Intuitions*, Oxford, Oxford University Press.
- SIMON A.-C., 2004, *La structuration prosodique du discours en français*, Berne, Peter Lang.
- SMITH S.W., NODA H. P., ANDREWS S. et JUCKER A. H., 2005, « Setting the stage: How speakers prepare listeners for the introduction of referents in dialogues and monologues », *Journal of Pragmatics*, 37, p. 1865-1895.
- SPROUSE J. et SCHÜTZE C.T., sous presse, « Grammar and the use of data », in AARTS B., BOWIE J. et POPOVA G., *Oxford Handbook of English Grammar*, Oxford, Oxford University Press.
- TALMY L., 2013, « Deixis and anaphora unified as ‘targeting’ », *International Journal of Cognitive Linguistics*, 4(2), p. 91-102.
- WILLEMS D., 1985, « La problématique des données et la place de l’exception en syntaxe contemporaine », *Langue Française*, 66, p. 86-98.