

La structuration informationnelle et la question du topic

Nous avons rendu compte, dans le chapitre précédent, de la question de la référence et des expressions référentielles, et avons abordé la question de leur usage dans le discours. Dans le large panorama des approches fonctionnalistes, dans lequel nous inscrivons notre travail, acquisition et emploi des expressions référentielles ne sauraient être séparés, et doivent être compris comme une (mise en) relation entre formes et fonctions. Le présent chapitre est en quelque sorte l'approfondissement du précédent. Nous allons cerner des relations entre formes et fonctions à un niveau plus global, au niveau de l'information et de sa structuration, et il sera montré que cette structuration concerne, au niveau paradigmatique, le choix des expressions référentielles ou les conditions de leur usage, et, au niveau syntagmatique, l'agencement linéaire des éléments au sein d'un énoncé. De fait, ces deux niveaux doivent être compris comme complémentaires, formes et agencement linéaire des expressions linguistiques participant, dans une relation complexe, à l'organisation discursive de l'information : ce dont on parle et ce que l'on en dit. Comme nous l'avons dit dans l'introduction, les travaux sur la structuration informationnelle et le statut attentionnel des référents concernant l'adulte ont nourri la réflexion actuelle sur l'acquisition du langage. Dans ce chapitre, nous allons donc d'une part présenter les travaux fondateurs sur la dimension pragmatique de la référence et des expressions référentielles. D'autre part, ce chapitre nous permettra de cerner de façon précise les différents aspects de la structuration informationnelle, de séparer notamment le statut attentionnel des référents, ou l'opposition *old/new*, de la dimension du topic-commentaire et de définir les notions centrales qui nous serviront dans nos analyses.

Ouvrons ce chapitre par une citation de « *Information Structure and sentence form* » (Lambrecht, 1994) :

« [...] the complex relationship between form and function is not arbitrary but motivated within the grammatical system of individual languages. While the form of constructions like the French *avoir-cleft* cannot be PREDICTED on the basis of communicative needs, this form can be shown to be MOTIVATED within the grammatical system of the language. Given the formal constraints of French grammar, and given the need to express a universal pragmatic category, it "makes sense" that the avoir-construction looks the way it does in spoken French. The notion of the PRAGMATIC MOTIVATION of grammatical form is one of the major theoretical concepts underlying the present study. » (1994 : 29)

Les clivées données en exemple dans cette citation ne sont pas l'objet de notre thèse, mais ce que dit Lambrecht là met le doigt sur l'articulation entre forme et fonction qui nous occupera. Lambrecht se garde de deux positions extrêmes, qui n'accorderaient pas assez ou, au contraire, trop d'importance aux besoins et fonctions communicatifs : Ainsi, fonctionnaliste, il se distancie d'une part du postulat de l'indépendance de la grammaire en cours dans les théories générativistes, en argumentant qu'il n'y aurait pas de forme (de grammaire, de syntaxe) sans fonction (besoins communicatifs). D'autre part, il rejette une conception de la grammaire qu'il qualifie de fonctionnalisme radical, et qui impliquerait que dans toutes les langues, les besoins communicatifs détermineraient des structures semblables, et que l'on puisse expliquer, voire prédire des formes à partir de leurs fonctions. Puisque les langues du monde contredisent aisément cette seconde hypothèse extrême, et puisque en outre les structures grammaticales possibles sont en nombre très limité comparé aux divers besoins communicatifs à exprimer, Lambrecht adopte un point de vue plus modéré : l'idée que les besoins communicatifs façonnent la forme des énoncés, selon un principe de motivation pragmatique, mais que ce principe opère dans le cadre des systèmes structurels propres à chaque langue. Selon Lambrecht, les processus qui, dans le développement d'une langue, font justement émerger ces structures, ne sont pas (encore) saisissables dans les théories linguistiques, mais il est néanmoins possible de retrouver la fonction pragmatique des constructions syntaxiques en s'intéressant à la structuration informationnelle. En ce qui nous concerne, nous allons nous poser la question de la relation entre forme et fonction non pas d'un point de vue de grammaticalisation diachronique, phylogénétique, mais du point de vue de l'ontogenèse, de l'acquisition de (quelques) formes et fonctions du langage.

Dans ce chapitre, nous proposons un survol des notions centrales relatives à la structuration informationnelle de l'énoncé essentiellement, avec des excursions vers la structuration des discours. Si la notion de thème/topic qui nous intéresse ne saurait se résumer à la seule dimension informationnelle (ni inversement), il s'agit néanmoins d'un facteur important. Au vu de la diversité des travaux sur la structuration informationnelle et de la complexité des notions, il nous a semblé nécessaire de les discuter dans une section dédiée, préalablement à l'examen de la notion de thème/topic. Cela nous a semblé utile également pour une autre raison : comme nous allons le montrer, dans nombre de travaux, information donnée et topic sont traités comme se définissant mutuellement. Or, et nous allons le montrer également, cette vue ne tient pas à l'examen des faits

linguistiques. Nous allons donc, dans ce chapitre, dresser d'abord le paysage général des travaux sur la structuration informationnelle. Face à la richesse de la littérature consacrée à ce domaine, nous serons contrainte d'opérer des choix, orientés par la thématique principale qui nous intéresse ici, le concept de thème/topic. Les sections 1 et 2 seront consacrées à une introduction générale des notions d'information et de structuration informationnelle, les sections suivantes en examineront des caractéristiques choisies. Ainsi, la section 3 présentera différentes approches et modèles de l'opposition entre information donnée/ information nouvelle (*given/new* ou *old/new*). La section 4 enfin sera consacrée à la notion centrale de thème/topic. Nous avertissons le lecteur que les notions de *comment* ou focus ne seront dans ce travail de thèse traitées pour eux-mêmes, mais par souci de place et de précision uniquement comme complémentaires de la notion de topic, là où des précisions s'imposent.

1. Avant de commencer : quelle définition de l'information ?

Ce chapitre traitera de la question de la structuration informationnelle dans les énoncés. Or, il est flagrant de constater que les travaux influents sur cette question proposent une théorie détaillée et argumentée sur l'idée de structuration, mais la notion de base, celle de l'information, n'est pas toujours explicitement thématifiée et théorisée. Halliday (1967) par exemple, à qui nous devons le terme d'*information structuring*, n'offre pas de définition explicite de la notion d'information elle-même. Ce que nous proposons alors dans cette section introductive est de retracer l'origine de la notion d'information dans les sciences du langage.

C'est avec le modèle de la communication formulé par Shannon & Weaver dans « *The mathematical theory of communication* » (1949) et conçu pour l'encodage de la télécommunication qu'a pu commencer, aussi dans la linguistique, la recherche d'un modèle formalisé de la communication et de l'information. Dans la théorie de l'information d'après Shannon & Weaver, l'information est une donnée essentiellement mathématique et probabiliste, mesurée par un logarithme. Weaver avertit qu'information dans cette optique n'est pas à confondre avec le contenu sémantique (*meaning*), et que

« information in communication theory relates not so much to what you *do* say, as to what you *could* say. That is, information is a measure of one's freedom of choice when one selects a message. If one is confronted with a very elementary situation where he has to choose one of two alternative messages, then it is arbitrarily said that the information, associated with this situation, is unity. Note that it is misleading (although often convenient) to say that one or the other message conveys unit information. The concept of information applies not to the individual messages (as the concept of meaning would), but rather to the situation as a whole, the unit information indicating that in this situation one has an amount of freedom of choice, in selecting a message, which it is convenient to regard as a standard or unit amount. » (Shannon & Weaver, 1949 : 9)

Un état ou événement constituent une information dans cette conception seulement lorsqu'ils ne sont pas entièrement prévisibles : « In the limiting case where one probability is unity (certainty) and all the others zero (impossibility), then H is zero (no uncertainty at all - no freedom of choice - no information) » (Shannon & Weaver, 1949 : 15). Cette mesure de l'incertitude H est appelée entropie. Pour expliquer ce raisonnement, il est souvent fait allusion à une source d'information qui ne peut produire qu'un seul symbole. Floridi (2017 : 2.1) fait appel au corbeau dans une nouvelle d'Edgar Allen Poe, qui répond « *nevermore* » à toute question qu'on lui pose. L'entropie ou l'incertitude de ce corbeau, compris en tant que source d'information, est alors zéro. Une « information » prévisible n'est donc pas information ; et plus grand est le choix possible ou l'incertitude, plus il y a d'information.

Le modèle de Shannon & Weaver a inspiré des travaux en sémiotique, dont notamment le célèbre schéma de la communication de Jakobson. Si l'information est conçue principalement comme un phénomène physique de quantité par Shannon & Weaver, les approches sémantiques et philosophiques de l'information s'intéressent au contraire aux conditions de la génération et de l'interprétation de l'information. Floridi (2017 : Chap. 4) résume ces investigations par les questions suivantes :

« “how can something count as information? and why?”, “how can something carry information about something else?”, “how can semantic information be generated and flow?”, “how is information related to error, truth and knowledge?”, “when is information useful?” »

La conception de l'information comme seulement ce qui est nouveau, ce qui peut être ajouté à la connaissance apparaissait déjà chez Peirce (1903 : 13) lorsqu'il écrivait « If you inform me of any truth, and I know it already, there is no information ». Dretske (1981 : 45) définit l'information comme un ajout à la connaissance (*knowledge*) : « A state of affairs contains information about X to just that extent to which a suitably placed observer could learn something about X by consulting it ».

Mais Dretske souligne également le fait que l'informativité d'un événement (linguistique ou autre) dépend de la situation de communication et des états mentaux des interlocuteurs, une dimension qui n'est pas prise en compte par la théorie de Shannon & Weaver. Il l'illustre par un exemple simplissime, mais très parlant :

« A glance at the calendar tells you the date only if you already know the day of the week. What one learns, or can learn, from a signal [...], and hence the information carried by that signal, depends in part on what one already knows about alternative possibilities. » (Dretske, 1981 : 43)

Cette notion sera centrale dans des approches davantage pragmatiques ou psychologiques de l'information. Ces théories ont recours, sous ce terme ou un autre, au concept de représentations mentales qu'un individu se fait du monde à un moment donné. Chafe (1994 : 27) par exemple dit que « every human being possesses a complex internal model of reality. Call it a worldview, call it a knowledge structure, this model is essential to the human way of coping with the world ». Bien que le concept central dans la théorie de Chafe ne soit pas l'information, mais ce qu'il considère comme relevant du concept de *consciousness*, les deux sont intimement liés :

« consciousness is the activation of only a small part of the experiencer's model of the surrounding world, not the model in its totality. This limited capacity of consciousness is reflected linguistically in the brief spurts of language that will be discussed as *intonation units* in chapter 5. Each such unit verbalizes a small amount of information which, it is plausible to suppose, is that part of the speaker's model of reality on which his or her consciousness is focused at that moment. » (Chafe, 1994 : 29)

Cette position se retrouve aussi chez Lambrecht, qui dédie un chapitre de sa monographie sur la structuration informationnelle à la notion de l'information : « One useful way of characterizing information is to say that by informing the hearer of some situation or state of affairs, the speaker influences the hearer's MENTAL REPRESENTATION of the world » (Lambrecht, 1994 : 43).

Lambrecht explique par la suite que si effectivement transmettre de l'information requiert que celle-ci soit nouvelle pour l'interlocuteur, il ne faut pas oublier que « the hearer's mind is not a blank sheet of paper on which new propositions are inscribed » (Lambrecht, 1994 : 45-46). Une telle approche ne réduit donc pas l'information aux seuls éléments nouveaux, mais la situe au contraire dans la relation établie par les interlocuteurs entre éléments connus et nouveaux, entre connaissances et ajouts relatifs à ces connaissances. Lambrecht rappelle que cette tension est déjà formulée par Strawson (1964) sous forme de deux principes complémentaires, le « Principle of the Presumption of Ignorance » et le « Principle of the Presumption of Knowledge ». Nous aurons l'occasion d'y revenir en section 4.3.3.1 ci-dessous.

Dans cette vision, il n'existe donc plus seulement de l'information nouvelle, mais également de l'information déjà connue ou donnée comme exprimé ici chez Chafe :

« given (or old) information is that knowledge which the speaker assumes to be in the consciousness of the addressee at the time of the utterance. So-called new information is what the speaker assumes he is introducing into the addressee's consciousness by what he says. » (Chafe, 1976 : 30)

La distinction entre information déjà connue et information nouvelle est cruciale pour toute théorie de la structuration de l'information, et nous reprenons ici une partie du passage de Dahl (1976), sur lequel se base Lambrecht, parce que cette citation introduit parfaitement les sections qui suivront :

« Let us consider one important use of declarative sentences, namely as means to influence the addressee's picture of the world. In such cases, the speaker assumes that the addressee has a certain picture - or model - of the world and he wants to change this model in some way. We might then identify THE OLD or THE GIVEN with the model that is taken as a point of departure for the speech act and THE NEW with the change or addition that is made in this model. » (Dahl, 1976 : 38)

Nous allons retrouver différents termes de cette citation dans les sections suivantes : l'information « donnée », l'information « nouvelle », « point de départ » ; et nous allons voir à quel point ces notions ont pu donner lieu à des confusions terminologiques. En effet, nous allons montrer que certaines approches ont mis l'accent sur les représentations mentales de référents comme donné ou nouveau, d'autres sur les relations de ces référents au sein d'une proposition, et que notamment « information donnée » et « point de départ » ont été conçus et interprétés de manière très différente. Prenons l'énoncé suivant : *Hier, ton père a appelé.* En fonction de la situation, l'interprétation de l'information sera très différente. Si dans le contexte précédent, l'on s'interrogeait sur les personnes qui ont téléphoné dans les derniers jours, on voudra certainement dire que *ton père* constitue une information nouvelle. Si au contraire, il était question du fait que le père n'a pas téléphoné aujourd'hui, c'est plutôt *hier* qui ajoute de l'information. Et si, dans un troisième scénario, l'on parlait de toute autre chose avant, l'énoncé tout entier pourrait être qualifié d'information nouvelle. Mais avec ces réflexions, le problème n'est pas encore épuisé, puisque si *ton père* peut, selon le contexte, être considéré comme information nouvelle dans un certain sens, la personne désignée par cette expression est très certainement déjà connue, encore dans un certain sens, de la personne à laquelle cet énoncé est adressé, indépendamment de la situation de communication. Chafe (1976 : 30), sur la base d'un exemple similaire, note que le lien traditionnellement fait entre information et *knowledge* a donné lieu à de grandes confusions dans la littérature en linguistique et en psychologie. Le problème serait alors de ne pas distinguer entre ce qui est connu de l'interlocuteur et de ce qui est actuellement au centre de sa conscience.

En ce qui concerne la définition de l'information appliquée à la linguistique et à l'énoncé, nous pouvons donc dire pour le moment qu'il y a au moins deux notions qui seront à distinguer. La première s'applique au statut des référents dans le modèle mental et discursif des locuteurs (ce qui est connu ou non, et ce dont on a déjà parlé récemment ou non). La deuxième concerne la manière dont ces référents sont présentés dans l'énoncé : si l'énoncé est conçu pour apporter une information nouvelle à propos du référent, ou si, au contraire, le référent est partie intégrante d'une information apportée à propos d'autre chose. Nous n'anticiperons pas plus ici sur cette question, qui a effectivement été source de définitions sensiblement différentes et de malentendus, car elle figurera comme fil rouge dans les sections 2 à 4, et est au cœur de la définition de la notion de topic.

Nous devons faire mention d'une autre question : la notion d'information est souvent, dans des approches philosophiques et sémantico-logiques, conditionnée par celle de vérité.¹¹ De la conception de l'information comme ajout à la connaissance s'ensuit pour Dretske qu'une information qui se révèle fautive n'en est pas une : « In this sense of the term, false information and mis-information are not kinds of information – any more than decoy ducks or rubber ducks are kinds of ducks. » (1981 : 45).¹²

Lambrecht au contraire plaide que les conditions de vérité ne s'appliquent pas à l'information conçue comme (modification des) représentations mentales : « I believe that in the linguistic analysis of information, hence in the grammatical domain of information structure, the logical concept of TRUTH has no place » (1994 : 44). Il étaye cette position par l'exemple suivant :

« If someone informs me that "The cat in the hat is back," my representation of the world is increased by one proposition, independently of whether what I'm being told is true. If later I find out that the proposition "The cat in the hat is back" was not true in the situation in which it was used, the representation of the cat being back may nevertheless linger in my mind. And this representation does not become false just because it does not correspond to the world as it is. It just becomes outdated. » (Lambrecht, 1994 : 45)

Le fait même que l'on puisse contrer une information dont on sait qu'elle est fautive (*il n'est pas vrai que X*) montre pour lui que la représentation mentale est indépendante des états de fait.

Une dernière remarque s'impose pour définir ce qu'est l'information dans une approche pragmatique du langage. Dretske écrit, dans l'introduction de son chapitre *Communication and Information*, que « a genuine theory of information would be a theory about the content of our messages, not a theory about the form in which this content is embodied » (1981 : 40). Or, pour revenir à notre citation de Lambrecht au tout début de ce chapitre, une des idées principales de la structuration informationnelle est « pragmatic motivation of grammatical form ». Celle-ci se manifeste aussi bien au niveau de la structuration linéaire des informations dans la production d'un énoncé (et de discours plus longs) qu'au niveau du choix des expressions référentielles. La question des expressions référentielles et des conditions pragmatiques de leur production a déjà été évoquée dans le CHAPITRE I, et nous aurons l'occasion d'approfondir cette question en présentant différents modèles du 'choix' référentiel. Le choix du matériel linguistique pour encoder un référent dépend de la représentation mentale de ce dernier que le locuteur peut supposer chez son interlocuteur. Si le locuteur assume qu'une entité n'est pas encore connue de l'interlocuteur ou que sa représentation mentale n'est pas actuellement activée chez ce dernier, il n'emploiera prototypiquement pas la même expression pour y référer que s'il peut supposer le référent comme déjà connu du locuteur et

¹¹ Valeurs de vérité ou *truth conditions*, notion déjà évoquée dans le CHAPITRE I sur la référence. Voir aussi section 4.3.3.1 ci-dessous sur les valeurs de vérité dans la théorie de la référence chez Strawson (1964).

¹² Pour une discussion de cette position et sa critique, voir Floridi (2017).

activé dans sa conscience : dans le premier cas, il dira par exemple *j'ai vu un chien*, alors que dans le deuxième cas, *je l'ai vu* sera suffisant.

Ce principe d'accessibilité est lié notamment aux travaux de Sperber & Wilson (1982, 1996 [1986]), dans leur élaboration des maximes conversationnelles de Grice (1975). Givón (voir notamment 1983) et Ariel (1988, 1990)¹³ développent la notion d'accessibilité et l'expriment explicitement comme une relation iconique entre forme et fonction : plus une information est prévisible, parce que déjà activée et au centre de la conscience des locuteurs, moins de matériel linguistique est nécessaire pour l'encoder, et inversement. Givón le formule ainsi : « The more disruptive, surprising, discontinuous or hard to process a topic is, the more *coding material* must be assigned to it » (1983 : 18). Mais il ne s'agit pas là d'un principe probabiliste de quantité comme dans la théorie de Shannon & Weaver, mais d'une relation pragmatique entre formes et fonctions. Cette question sous-tend en partie les développements présentés en section 3 et notamment les différentes hiérarchies et échelles d'états de représentations mentales et leur expression linguistique.¹⁴

Ce qui nous occupera par la suite pour aborder la notion de topic est la question de la structuration informationnelle de l'énoncé. En cela, la tension entre information donnée et information nouvelle, point de départ et ajout mentionnés dans la citation de Dahl (1976) ci-dessus peut être comprise comme principe fondamental. La section suivante s'attachera à expliciter cette relation et à montrer l'intrication entre forme et fonction à ce niveau. Il deviendra clair que, dans une approche pragmatique du langage, l'information ne peut être conçue indépendamment du code qui sert à la transmettre.

2. Qu'est-ce que la Structuration Informationnelle ?

En guise d'entrée en matière, commençons par un exemple simple :

Exemple II-1

- (a) J'aime les cerises.
- (b) Les cerises, j'aime.

L'intérêt pour la structuration informationnelle prend sa source dans la nécessité d'expliquer la différence d'interprétation pragmatique et d'emploi entre des énoncés comme dans l'exemple ci-dessus. Ce qui les distingue n'est pas d'ordre sémantique : le contenu propositionnel est le même dans les deux énoncés, les rôles sémantiques des actants du verbe identiques, *je* exprimant le rôle d'expérimenteur du sentiment exprimé par le verbe, *les cerises* en exprimant le patient. Les fonctions syntaxiques ne changent pas non plus entre Exemple II-1 (a) et (b) : *je* est sujet, *les cerises* objet du

¹³ Voir sections 3.2.3 et 4.3.4 pour une présentation de leurs modèles respectifs.

¹⁴ Voir aussi l'échelle de Givón (1983) en section 4.3.4.

verbe *aimer*. La différence tient dans l'ordre des mots. Mais qu'est ce qui motive le locuteur de produire soit *je*, soit *les cerises* en position initiale de l'énoncé ?

L'organisation interne d'un énoncé ne se résume pas aux seules dimensions sémantiques et syntaxiques. Pour communiquer de façon efficace, le locuteur doit orienter son énoncé selon ses propres buts communicationnels, ainsi que selon l'état de connaissances (supposé) de l'interlocuteur (Musan, 2010 : 1). L'ordre des mots, ainsi que l'intonation et des constructions syntaxiques spécialisées sont parmi les outils dont disposent les langues pour organiser les énoncés selon ces principes. Pour le résumer avec Marandin (2006),

« on postule un niveau d'organisation grammatical, qu'on appelle *structure informationnelle* pour rendre compte du fait que plusieurs énoncés qui ont le même contenu sémantique présentent une forme distincte, soit en termes de réalisation prosodique, d'ordre des mots, de type de construction ou de marquage des constituants. » (Marandin, 2006)

Lambrecht (1994) insiste, à juste titre, sur le fait que l'information fournie lors de la production d'un énoncé est à distinguer du contenu propositionnel de cet énoncé. Pour illustrer ce point de façon encore plus claire, prenons l'exemple suivant, où l'ordre des mots ne varie pas :

Exemple II-2 – exemple construit

	<u>Énoncé</u>	<u>Contexte dans lequel l'énoncé est acceptable</u>
(a)	Marie mange des cerises.	Quoi de neuf ?
(b)	MARIE mange des cerises.	Qui mange des cerises ?
(c)	Marie MANGE des cerises.	Est-ce que Marie vend des cerises ?
(d)	Marie mange DES CERISES.	Que mange Marie ?

Dans les énoncés (a)-(d) dans la colonne de gauche, le contenu propositionnel reste de nouveau le même, mais leur intonation est différente : les constituants en capitales correspondent à une accentuation intonative. Ces différences d'intonation induisent une lecture différente pour chaque énoncé : en effet, la question à laquelle pourrait répondre chaque énoncé n'est pas la même (cf. la colonne de droite).¹⁵

L'intérêt pour ces questions, pour la relation entre communication et forme linguistique, n'est pas récent. De fait, il trouve son origine dans les réflexions logico-philosophiques d'Aristote (*Les Catégories* et *Sur l'interprétation* dans l'*Organon* (2014) notamment) et de Platon (dans le *Cratyle* et le *Sophiste* notamment ; (Platon, 2008)). Dans les notions de *onoma* et *rhêma* chez Platon puis Aristote, selon les traditions traduites par nom et verbe, ou par sujet et prédicat, se trouve déjà l'idée d'une

¹⁵ Cet exemple représente les caractéristiques prosodiques des énoncés d'une façon très simplifiée, pour des besoins de présentation. En réalité, un énoncé comme (a) a très peu de chances d'être produit. Dans le cas d'un agencement prosodique neutre de l'énoncé, l'accent principal de l'énoncé tombera sur la fin du dernier groupe intonatif. De ce fait, l'énoncé (d) pourra répondre à la fois aux questions (a) et (d), c'est-à-dire qu'il peut être interprété comme ayant un focus large, voire étant entièrement en focus, répondant alors à la question *Quoi de neuf?*, ou bien être en focus étroit, répondant à la question *Que mange Marie ?*. La prosodie du français et de l'allemand sera abordée dans le CHAPITRE III. Une première précision sera donnée lorsque nous reprendrons cet exemple dans la section 3.1.1 ci-dessous.

bipartition de la proposition, en une entité dont on parle, et une autre qu'on affirme (ou nie) de la première (mais cf. Basset, 2004 pour plus de précision).

« Un nom est donc un vocable signifiant par convention, sans référence à un temps [...]. Un rhème est ce qui ajoute une signification temporelle et dont aucune partie ne signifie séparément ; et il est toujours signe de choses dites d'une autre. Je dis qu'il « ajoute une signification temporelle » : par exemple santé (hugieia) est un nom, est-en-bonne-santé (hugiainei) est un rhème car il signifie en plus que c'est un attribut maintenant. Et je dis qu'un rhème est toujours [...] signe d'attributs : par exemple, de choses qu'on dit d'un sujet. » (Aristote, 2014 : 68 (*Sur l'Interprétation*))

La visée de ces réflexions est de déterminer la logique sous-jacente au langage, le lien entre pensée et langage, et il s'agit de déterminer dans quelle mesure le langage est fidèle à la réalité, et sous quelles conditions une proposition sera vraie ou fausse. Ces préoccupations donneront lieu à la tradition logique de la philosophie du langage, (de la *Grammaire générale et raisonnée* (Arnauld & Lancelot, 1660 ; dite *Grammaire de Port Royal*) ou la *Logique de Port Royal* (Arnauld & Nicole, 1874 [1662]) jusqu'aux travaux de Frege (1892) et de Russell (1905)). La bipartition de la phrase en sujet et prédicat dérive de ces premiers travaux classiques, esquissée chez Platon et Aristote, entérinée comme principe grammatical dans la *Grammaire de Port Royal* (sous les termes « sujet » et « attribut » (Arnauld & Lancelot, 1660 : 29)) puis dans la *Logique de Port Royal* (où l'attribut est aussi appelé « affirmation » ou « *praedicatum* » (Arnauld & Nicole, 1874 [1662] : 105)).

Au 19^{ième} siècle, la relation entre grammaire et psychologie motivera des nouvelles réflexions sur ces mêmes phénomènes, et permettra d'introduire la distinction entre un niveau grammatical/syntaxique et un niveau dit psychologique. G. von der Gabelentz (1891 : 345) reconnaît ainsi que, si les langues permettent de dire la même chose de différentes façons, cela correspondrait, d'un point de vue qu'il appelle « *geschäftlich* » (transactionnel), à un surplus, une variété superflue de moyens d'expression. Il donne pour exemple des variations dans l'ordre des mots pour un même contenu propositionnel, ou encore des expressions prédicatives versus attributives : « *hoher Berg* » ('montagne haute') versus « *der Berg ist hoch* » ('la montagne est haute'), où l'idée évoquée resterait la même (« *erweckt dieselbe Gesamtvorstellung* », (1891 : 346)). Selon Gabelentz alors, il doit exister une force qui motive ces variations (« *Formungstrieb* », (1891 : 347)), et qui s'oppose à l'inertie qui élève en règle ce qui est habituel (« *Der menschliche Geist lässt wohl gern in seiner Trägheit das Gewöhnliche zur bindenden Regel werden.* », (1891 : 346)).¹⁶ Gabelentz poursuit en examinant les variations dans l'ordre des mots évoqués plus haut, et c'est ainsi qu'il introduit les notions de 'sujet

¹⁶ Cette conception de Gabelentz traduit une certaine reconnaissance des besoins communicatifs qui conditionnent les formes que prend la langue. Mais en même temps, Gabelentz ne semble pas se défaire complètement de l'idée d'une structure canonique, la règle. Les travaux fonctionnalistes pousseront ces réflexions plus loin, considérant la langue comme produit des usages qu'en font les locuteurs (cf. notre CHAPITRE V).

et prédicat psychologiques',¹⁷ qui peuvent différer des sujet et prédicat grammaticaux, et qu'il définit ainsi :

« Welches aber ist dieser Antrieb? Das Bild des Ganzen schwebt mir vor; die Theile halte ich in Händen, um sie nachschaffend aufzubauen. Was bestimmt mich, erst diesen aufzustellen, dann den, dann jenen? Offenbar ist es dies, dass ich erst dasjenige nenne, was mein Denken anregt, worüber ich nachdenke, mein psychologisches Subjekt, und dann das, was ich darüber denke, mein psychologisches Prädikat [...]. » (Gabelentz, 1891 : 353-354)¹⁸

À la suite de Gabelentz, Hermann Paul (1920 [1880]), dans les « *Principien der Sprachgeschichte* » (1920 [1880]) affinera l'idée d'un fondement psychologique de la langue : ainsi développera-t-il plus explicitement l'idée que chaque catégorie grammaticale aurait pour origine une catégorie psychologique. Dès la deuxième édition de son ouvrage (1886 [1880]), il reprend à son compte les notions de sujet et prédicat psychologiques.¹⁹ À côté de l'agencement linéaire des mots au sein de l'énoncé déjà observé par Gabelentz, Paul ajoute deux autres traits caractéristiques de la structuration informationnelle. Il s'agit de l'intonation (Paul, 1920 [1880] : 123, 283-284) et, de façon plus marginale, l'idée qu'à cette structuration correspond d'un côté une information déjà connue, de l'autre côté une information nouvelle et plus importante pour la visée communicative : « Im isolierten Satze ist das psychologische Prädikat als das bedeutsamere, das neu hinzutretende stets das stärker betonte Element » (1920 [1880] : 126).

L'introduction du terme moderne d'« *Information Structure* » est attribuée à Halliday (1967), qui lie à la notion de 'thème' et le décrit comme étant le niveau où des choix syntaxiques sont opérés en fonction d'éléments en tant que composantes d'un message (opposé donc au contenu sémantique), du lien qui est fait avec le discours antérieur et finalement l'organisation interne de la proposition en fonction de ces critères :

« Theme is concerned with the information structure of the clause; with the status of the elements not as participants in extralinguistic processes but as components of a message; with the relation of what is being said to what has

¹⁷ Il convient de noter que c'est déjà dans un court article de 1869 que Gabelentz introduit ces notions. Nous avons préféré les citer de son ouvrage principal afin de montrer la cohérence globale de cette orientation psychologique du langage et de la langue. Il est par ailleurs intéressant qu'en 1869, il formule les notions de sujet et prédicat psychologiques non pas du point de vue du locuteur, comme dans l'ouvrage de 1891, mais du point de vue de l'interlocuteur : « ce à quoi je veux faire penser l'interlocuteur » et « ce qu'il doit en penser » (Gabelentz, 1869 : 378; notre traduction).

¹⁸ « Quelle est donc la motivation ? L'image de l'ensemble est devant moi ; je tiens les parts dans mes mains pour le reconstituer. Qu'est ce qui me motive de poser d'abord celui-ci, puis celui-là, puis tel autre ? Apparemment est-ce ainsi, que je nomme d'abord ce qui stimule ma pensée, ce à quoi je réfléchis, mon sujet psychologique, et ensuite ce que j'en pense, mon prédicat psychologique [...]. » (Gabelentz, 1891 : 353-354; notre traduction)

¹⁹ „Jede grammatische Kategorie erzeugt sich auf Grundlage einer psychologischen.“ (Paul, 1920 [1880] : 263), ainsi que „Jeder Satz besteht demnach aus mindestens zwei Elementen. Diese Elemente verhalten sich zu einander nicht gleich, sondern sind ihrer Funktion nach differenziert. Man bezeichnet sie als Subjekt und Prädikat. Diese grammatischen Kategorien beruhen auf einem psychologischen Verhältnis. Zwar müssen wir unterscheiden zwischen psychologischem und grammatischem Subjekt, respektive Prädikat, [...] da beides nicht immer zusammenfällt [...]. Aber darum ist doch das grammatische Verhältnis nur auf Grundlage des psychologischen aufbaut.“ (Paul, 1920 [1880] : 124)

gone before in the discourse, and its internal organization into an act of communication [...]. » (Halliday, 1967 : 199)

Prince (1981) précise l'intérêt linguistique de la structuration informationnelle dans les termes suivants :

« We may now word the basic problem as follows. From the point of view of a speaker/writer, what kinds of assumptions about the hearer/reader have a bearing on the form of the text being produced, where that form is not uniquely determined by the "objective" information that the speaker/writer is attempting to convey? From the point of view of the hearer/reader, what inferences will s/he draw on the basis of the particular form chosen? We are, therefore, NOT concerned with what one individual may know or hypothesize about another individual's belief-state EXCEPT insofar as that knowledge and those hypotheses affect the forms and understanding of LINGUISTIC productions. » (1981 : 233)

Lambrecht (1994 : 3), en reprenant cette citation de Prince, insiste sur le fait que la structuration informationnelle est intéressante pour le linguiste en tant qu'elle est reflétée sous des traits morpho-syntaxiques ou prosodiques dans la structure grammaticale. Il précise que pour lui, la structuration informationnelle fait partie de la grammaire : « I take information structure to be a component of GRAMMAR, more specifically of SENTENCE GRAMMAR, i.e. I take it to be a determining factor in the formal structuring of sentences. » (1994 : 3).

L'idée que la structuration de l'énoncé réponde non seulement aux exigences structurelles de la langue, mais également aux besoins communicatifs des locuteurs se retrouve dans différentes notions centrales, travaillées, explicitement ou non, sous le grand thème de la structuration informationnelle. Cet angle d'approche sous-tend explicitement les théories linguistiques fonctionnalistes, mais d'autres approches, générativistes (cf. notamment section 4.3.1 ci-dessous) par exemple, s'intéresseront à la structuration informationnelle sans pour autant envisager que ce niveau d'organisation ait une influence sur les structures syntaxiques. Certaines traditions se sont essentiellement consacrées aux notions de thème/rhème ou topic/focus (sous l'un de ces noms ou sous d'autres, cf. section 4 ci-dessous), d'autres se sont davantage intéressés à la distinction information ancienne/ information nouvelle (cf. section 3), d'autres encore ont considéré ces phénomènes à partir des notions de *common ground* ou *shared knowledge*, bien qu'il ne s'agisse pas exactement du même angle (cf. section 3.1.2 spécifiquement). Ces notions ont pu être traitées pour certaines comme équivalentes, concordantes, ou encore complémentaires, de sorte qu'en fournir un tableau synthétique n'est pas tâche aisée. Notamment les dimensions de l'information ancienne/nouvelle et celle de thème/topic apparaissent comme liées de façon très complexe. Dans ce qui suit, nous allons exposer les notions et approches qui nous paraissent incontournables pour notre travail, que nous tenons pour représentatives sans pouvoir espérer les traiter de manière exhaustive.

3. L'asymétrie informationnelle

Nous choisissons d'entrer dans cette section avec deux observations qui paraissent parfaitement banales et qui font écho à ce que nous avons introduit en section 1 ci-dessus : le fait qu'en parlant, premièrement, nous transmettons, entre autres et le plus souvent, de l'information. Deuxièmement, et c'est là le point crucial, que cette information est présentée de façon asymétrique, avec certaines unités plus « anciennes » que d'autres. Prince (1981) le formule ainsi :

« One presumably universal feature of natural language is that the objective information conveyed is not conveyed on a single plane. That is, there is an INFORMATIONAL ASYMMETRY in that some units seem to convey or represent "older" information than others. » (Prince, 1981 : 224)

Nous avons montré en section 2 que cette observation était déjà présente chez Paul, lorsqu'il estime que le prédicat psychologique correspond à l'information nouvelle : « das neu hinzutretende [...] Element » (1920 [1880] : 126).²⁰ La reconnaissance d'une association entre sujet/prédicat psychologique et information ancienne/nouvelle n'est donc pas récente, mais cette association constitue plus une tendance de l'encodage linguistique qu'un principe universel et absolu, comme notre revue de la littérature montrera. C'est pour cette raison que nous consacrons des sections distinctes à ces phénomènes. Nous allons, dans la suite de la présente section, présenter différentes conceptions de cette dimension de la structuration informationnelle qui concerne le statut de l'information comme plus ou moins 'ancienne' ou 'nouvelle'. Si nous employons le terme d'information ici, il convient de noter qu'à partir des années 80, il semble s'opérer un certain glissement concernant cette notion : en effet, nombre de travaux s'intéresseront davantage aux questions de la référence, de l'antécédent d'une expression référentielle et à l'anaphore. Certains auteurs n'utiliseront donc plus le terme d'information, mais parleront davantage de statut attentionnel et d'accessibilité. Malgré ce changement terminologique, la question du statut de l'information comme ancienne/nouvelle, *old/new* ou *given/new* restera centrale aux travaux sur la référence.

3.1 Différentes définitions de 'old/new'

Pour commencer l'investigation, l'on peut recenser différents couples de termes dans la littérature qui semblent décrire le même phénomène. Toutefois, nous allons voir que les définitions données à ces termes ne coïncident pas nécessairement.

²⁰ « l'élément nouveau qui vient s'ajouter » (*notre traduction*)

ancien/nouveau	old/new
donné/nouveau	given/new
connu/nouveau	known/new
moins informatif/plus informatif	

L'intérêt du linguiste pour cette organisation de l'information réside dans l'observation qu'elle se manifeste dans l'organisation séquentielle de l'énoncé, c'est-à-dire dans l'ordre des mots (cf. section 2 ci-dessus), ainsi que dans l'intonation et le choix des expressions référentielles :

« The general thing to say is that given concepts are spoken with an attenuated pronunciation [sic]. The attenuation involves, at the very least, weak stress. Typically, though not always, it also involves either pronominalization or omission from verbalization altogether (the maximum degree of attenuation). »
(Chafe, 1987 : 26)

Pour la présentation des différentes définitions avec lesquelles l'on a tenté de saisir ce phénomène, nous allons suivre la synthèse opérée par Prince (1981), qui distingue dans la littérature, sous ces termes ou sous d'autres, trois niveaux de *givenness* :

« In fact, if one considers the definitions that have been presented, one discovers that there is not one notion involved but (at least) three. Of all the linguists who have used these terms or their synonyms, those that are perhaps the most strongly associated with the notion are Chafe, Clark and Haviland, Halliday, and Kuno . It is rather surprising, then, that, when their discussions on the subject are closely examined and compared, it becomes evident that no two of them mean quite the same thing, and that, in some cases, the differences are quite large. »
(Prince, 1981 : 225)

Un premier groupe de définitions est résumé sous les termes de prévisible/récupérable (« predicatability/recoverability » (1981 : 226)), un deuxième sous l'étiquette de savoir partagé (« shared knowledge » (1981 : 230)), et un troisième sous l'étiquette de saillance (« saliency » (1981 : 228)).²¹

3.1.1 L'information présentée comme 'prévisible/récupérable'

Prince (1981) associe les positions de Kuno (1972, 1978, 1981)²² et de Halliday (1967) ainsi que Halliday & Hasan (1976) avec une conception de l'information comme plus ou moins prévisible ou récupérable. Bien qu'elle note que cette notion n'est pas définie de la même façon chez ces auteurs, le point commun est résumé par Prince comme suit, en tant qu'un premier principe définitoire de *givenness* :

²¹ Nous nous permettons d'invertir l'ordre de présentation par rapport à la synthèse de Prince (1981), qui présente la notion de *shared knowledge* en dernier : il nous a semblé préférable de représenter la chronologie d'apparition des notions et les traces d'affinement qu'elles ont connues à travers le temps.

²² Prince (1981) mentionne Kuno (1979), identifié comme « unpublished » dans la bibliographie. Ce travail a été publié dans Bedell, Kobayashi, & Muraki (1981) et est répertorié dans notre bibliographie en tant que Kuno (1981).

« The speaker assumes that the hearer CAN PREDICT OR COULD HAVE PREDICTED that a PARTICULAR LINGUISTIC ITEM will or would occur in a particular position WITHIN A SENTENCE. » (Prince, 1981 : 206)

La définition de Prince concerne donc une forme linguistique particulière dans une position syntaxique particulière. L'auteure ne précise pas explicitement ce qu'elle entend par là, mais la suite de l'article ainsi que les exemples donnés suggèrent qu'elle pense particulièrement aux phénomènes de pronominalisation et aux anaphores zéro pour les « particular linguistic items » ainsi que, pour la position, à des préférences présumées dans l'ordre d'apparition d'éléments dans un énoncé, exprimé par le *From-Old-to-New Principle* de Kuno (1981), c'est-à-dire l'hypothèse d'une tendance à exprimer de l'information ancienne avant l'information nouvelle. Pour Halliday, l'information donnée aura également tendance à se trouver en début d'un énoncé, ce qu'il schématise sous forme de la formule « (given followed by) new » (1967 : 205).²³

Regardons maintenant dans le détail les positions de Kuno, puis de Halliday.

Pour Kuno (1972), il est nécessaire de distinguer entre le statut d'un référent comme *old/new* et les relations sémantiques que cet élément entretient avec le reste de la phrase. Dans le premier cas, il s'agit de relations anaphoriques, et il entend par là un élément mentionné dans le discours antérieur, ou dont le référent est « uniquely identifiable » (1972 : 272). Au deuxième cas, Kuno réserve le couplage terminologique *old/new* dans le sens où la relation entre un élément et le reste de la phrase est plus ou moins prévisible. Il donne un exemple simple pour illustrer cette distinction :

Exemple II-3 - Kuno (1972 : 272)

- (a) Among John, Mary and Tom, who is the oldest?
(b) Tó^hm is the oldest.²⁴

Dans la réponse (b), l'élément *Tom* est anaphorique, mais représente en même temps l'information nouvelle en tant qu'élément à mettre en relation avec l'information ancienne, prévisible, que constitue le reste de la phrase, *is the oldest*. Kuno résume cette distinction ainsi :

« Thus, it is clear that we have to distinguish between two different concepts related to old and new information: the concept applied to lexical items, on one hand, and the concept applied to the particular semantic relations which lexical items enter into in the given sentence. I will use the terms "anaphoric/nonanaphoric" strictly for the former sense, and the terms "old/new information" strictly for the latter sense. In [I-3(a)], the lexical item Tom is anaphoric because it must have been previously mentioned in the discourse and/or its referent is uniquely identifiable. On the other hand, it conveys new information in [I-3(b)] in that it is the subject of *is the oldest*. »(Kuno, 1972 : 272)

²³ mais voir la section 4.3.2 ci-dessous pour des précisions

²⁴ L'accent graphique sur *Tó^hm* représente une accentuation tonique de cet élément là.

Kuno précise dans une note de bas de page (ibid.) que son opposition *old/new* correspond à présupposition/focus dans d'autres travaux. Sa dimension anaphorique, quant à elle, couvre les référents mentionnés préalablement ou autrement connus, entrés dans la mémoire à long terme (« permanent registry » (1972 : 271)).

Comme le note Prince (1981), Halliday (1967), à la différence de Kuno, base sa définition en premier lieu sur des critères intonatifs :

« Information focus is one kind of emphasis, that whereby the speaker marks out a part (which may be the whole) of a message block as that which he wishes to be interpreted as informative. What is focal is 'new' information; not in the sense that it cannot have been previously mentioned, although it is often the case that it has not been, but in the sense that the speaker presents it as not being recoverable from the preceding discourse. » (Halliday, 1967 : 204)

Pour mieux illustrer ce point, nous pouvons maintenant reprendre et affiner l'Exemple II-2 de la section 2 ci-dessus :

Exemple II-4 – exemple construit

- | | | |
|-----|--------------------------|---|
| (a) | maRIE mange des cerises. | Qui mange des cerises aujourd'hui? |
| (b) | Marie MANGE des cerises. | Est-ce que Marie vend des cerises aujourd'hui ? |
| (d) | Marie mange des ceRISES. | Quoi de neuf ?/ Que mange Marie aujourd'hui ? |

Halliday distingue entre ce qu'il appelle le focus marqué et le focus non marqué. Dans le cas de focus non-marqué, l'accentuation tombe sur la dernière syllabe du groupe intonatif, comme en (d). Dans ce cas, l'élément accentué est marqué comme étant nouveau, mais le statut du reste de l'énoncé n'est pas spécifié : le domaine du focus pourrait se restreindre au seul constituant *des cerises* (et correspondre donc à la question *Que mange Marie aujourd'hui*), mais il pourrait également s'étendre à l'ensemble de l'énoncé, auquel cas l'énoncé entier sera présenté comme nouveau, non prévisible, répondant à une question comme *Quoi de neuf*.

3.1.2 L'information dans le cadre du 'savoir partagé'

La distinction entre de l'information *given* et *new* chez H. H. Clark & Haviland (1977) est définie en fonction du savoir ou des connaissances que le locuteur attribue à l'interlocuteur :

« (1) Information the speaker considers given – information he believes the listener already knows and accepts as true; and (2) information the speaker considers new – information he believes the listener does not yet know. » (H. H. Clark & Haviland, 1977 : 3)

Locuteur et interlocuteur adhèreraient alors à un contrat implicite, le « given-new-contract » ou « given-new-strategie » (H. H. Clark & Haviland, 1974, 1977; Haviland & Clark, 1974), qui régit la production et l'interprétation des énoncés en fonction de ces assumptions. Les auteurs font

explicitement référence aux maximes conversationnelles de Grice (1975) pour expliquer comment l'interlocuteur interprète l'information présentée soit comme *given*, soit comme *new* de la part du locuteur, en la comparant à ce qu'il sait effectivement.

Dans Clark & Haviland (1977), le *given-new-contract* est formulé ainsi :

« Try to construct the given and the new information of each utterance in context (a) so that the listener is able to compute from memory the unique antecedent that was intended for the given information, and (b) so that he will not already have the new Information attached to that antecedent. » (1977 : 9)

Dans les différentes publications citées ici, les auteurs expliquent que l'accès à l'antécédent visé peut se faire de diverses manières ; de façon directe lors d'une mention précédente, ou bien de façon indirecte, dont un exemple est donné ci-dessous :

Exemple II-5 - (H. H. Clark & Haviland, 1977 : 21)

- (a) Horace got some beer out of the car. The beer was warm.
- (b) Horace got some picnic supplies out of the car. The beer was warm

Dans la séquence (a), l'interlocuteur peut trouver un antécédent direct pour l'expression *the beer* du deuxième énoncé : les bières sont mentionnées dans le premier. Dans la séquence (b) en revanche, il n'y a pas d'antécédent direct, et l'interlocuteur doit procéder à des implicatures et comprendre que parmi les *picnic supplies*, il devait y avoir de la bière, et que c'est de ces bières-là qu'il est question dans le second énoncé. Ces procédures-là sont aussi appelées *bridging* (cf. aussi H. H. Clark, 1975).

Clark & Haviland (1974) livrent une description détaillée des processus cognitifs menant à l'interprétation des énoncés, en fonction de la *given-new-strategy*. Pour les auteurs, la deuxième phrase dans l'Exemple II-6 (b) est inacceptable en raison d'une violation justement de la *given-new-strategy* :

Exemple II-6 – tiré de Clark & Haviland (1974 : 108)

- (a) Someone was following Mary. It was John who was following Mary.
- (b) John was following someone. It was John who was following Mary.

En effet, selon les auteurs, la deuxième phrase dans (b), *It was John who was following Mary* serait inacceptable parce qu'elle présente le fait que c'était John qui suivait quelqu'un comme information nouvelle (alors que cette information était connue après avoir entendu la première phrase de (b)), et, à l'inverse, le fait que Mary était suivie par quelqu'un est présenté comme *given*, bien que l'interlocuteur n'en pût pas encore en avoir connaissance ayant entendu la première phrase de (b).

Dans cette interprétation, les auteurs semblent amalgamer les dimensions que Kuno justement distingue, ce qui est *given* référentiellement, et ce qui est *given* relationnellement.²⁵ Cette simplification est à la base de l'amalgame entre thème/topic et information donnée, comme nous le

²⁵ (cf. aussi J. K. Gundel, 1999)

montrons plus explicitement en section 4 ci-dessous, et la séparation de ces deux notions sera aussi à la base de la définition de topic adoptée dans cette thèse.

3.1.3 L'information saisie en termes de 'saillance'

L'acceptation de *givenness* dans le sens de 'saillance' est associée, dans l'article de Prince (1981) aux travaux de Wallace L. Chafe (tels que présentés dans Chafe (1976)). L'apport central de Chafe aux questions de la structuration informationnelle, et au langage en général, est la notion de *consciousness* (notamment Chafe, 1974). Cette notion, centrale dans ses travaux depuis les années 70, est synthétisée dans la monographie *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing* (Chafe, 1994). La relation entre *consciousness* et structuration informationnelle est formulée ainsi :

« Given (or old) information is that knowledge which the speaker assumes to be in the consciousness of the addressee at the time of the utterance. So-called new information is what the speaker assumes he is introducing into the addressee's consciousness by what he says. » (Chafe, 1976 : 30)

Par rapport à la conception en termes de *shared knowledge*, la proposition de Chafe présente l'avantage d'être dynamique, d'envisager que des informations puissent se trouver dans la conscience d'un locuteur à un moment donné, et faire place à d'autres au cours de l'interaction. Au départ, cette notion est conçue comme binaire, ce qui a comme conséquence que des entités présumées connues seront aussi 'neuves' que des entités non connues, voir l'exemple suivant donné par Prince (1981) :

Exemple II-7 - (Prince, 1981 : 229)

- (a) I saw your father yesterday.
- (b) I saw a two-headed man yesterday.

De même, à la différence de Clark & Haviland mentionné plus haut, Chafe (1976) considère comme *given* seulement les entités dont le locuteur peut présumer qu'ils se trouvent dans la conscience de l'interlocuteur au moment de produire un énoncé. L'auteur précise qu'un élément peut se trouver dans la conscience des locuteurs de deux façons : il peut avoir été mentionné dans le discours antérieur, ou le locuteur peut assumer qu'il partage avec l'interlocuteur la même perception du contexte extralinguistique (Exemple II-8 (a) ci-dessous). Dans le cas d'un déclencheur linguistique du statut *given*, après la mention d'un élément, les mentions subséquentes le traiteront comme *given* (Exemple II-8 (b) ci-dessous). L'Exemple II-8 (c) ci-dessous montre que le statut de *given* peut s'étendre également à d'autres exemplaires d'une même catégorie.

Exemple II-8 : Exemples tirés de Chafe (1976 : 31-32)

	Contexte linguistique ou extralinguistique comme déclencheur du statut ‘given’ :	Présentation d’un élément comme ‘given’ (en gras)
(a)	Contexte extralinguistique : (Le locuteur voit que l’interlocuteur regarde un tableau accroché au mur)	<i>I bought it last week.</i>
(b)	Contexte linguistique : <i>I’d like to show you a painting.</i>	<i>I bought it last week.</i>
(c)	Contexte linguistique : <i>There was a small earthquake last night.</i>	<i>I felt one last year at about this same time.</i>

En revanche, un élément que l’interlocuteur peut identifier seulement en recourant à des inférences plus complexes ne sera pas interprété comme *given* par Chafe. Ainsi, *the beer* de l’exemple xx ci-dessous est *given* dans la conception de Chafe en (a), mais pas en (b).

Exemple II-9 – adapté de Chafe (1976 : 41)²⁶

- (a) We got some beer out of the trunk. **The beer** was warm.
- (b) We checked the picnic supplies. **The beer** was warm.

Ceci ne veut pas dire que Chafe ignore les conditions dans lesquelles un référent peut être introduit sous forme définie, mais il ne saisit cette relation justement pas en termes de *givenness*, mais de *definiteness* ou *identifiability*. Il reconnaît donc le comportement particulier de ce qui a été appelé ailleurs ‘anaphore associative’ (en anglais, *bridging* (cf. i.a. H. H. Clark, 1975; H. H. Clark & Haviland, 1977) ou *inferrable* (cf. Prince, 1981), souvent illustré par l’exemple suivant : *Nous arrivâmes dans un village. L’église était fermée.* (voir par exemple Apothéloz, 1995 : 40 sqq.), où la mention de *l’église* présente cet élément comme connu ou identifiable. Selon Chafe (1976), ceci est possible parce que la mention d’un premier référent particulier (ici donc, les *picnic supplies* ou le *village* respectivement) permet à l’interlocuteur d’identifier sans ambiguïté des bières ou une église particulières, mais il ne s’ensuit pas, pour Chafe, une différence avec des éléments non identifiables en ce qui concerne l’intégration d’un nouvel élément dans la conscience du locuteur ; des référents connus, non connus et identifiables par inférence seraient tous introduits comme *new* dans la conscience de l’interlocuteur.

3.2 Élargissement de la conception binaire ‘*given/new*’ à des modèles plus complexes

Dans un souci d’affinement de la bipartition initiale des réflexions sur l’information donnée et nouvelle, différents auteurs ont proposé des modèles plus complexes, que nous résumerons dans la présente section, comme les états d’activation de Chafe (1987), la taxinomie et l’échelle de familiarité présumée (*Assumed Familiarity*) de Prince (1981), l’échelle d’accessibilité (*Accessibility scale*) d’Ariel (1988, 1990) et la *givenness hierarchy* de Gundel, Hedberg & Zacharski (1993). D’autres échelles, de premier abord très similaires, sont davantage rattachées à la notion de topic (*topic accessibility scale*

²⁶ à l’origine, l’exemple provient de Haviland & Clark (1974)

de Givón (1983) ; *topic acceptability scale* de Lambrecht (1994), et seront en conséquence traités en section 4.

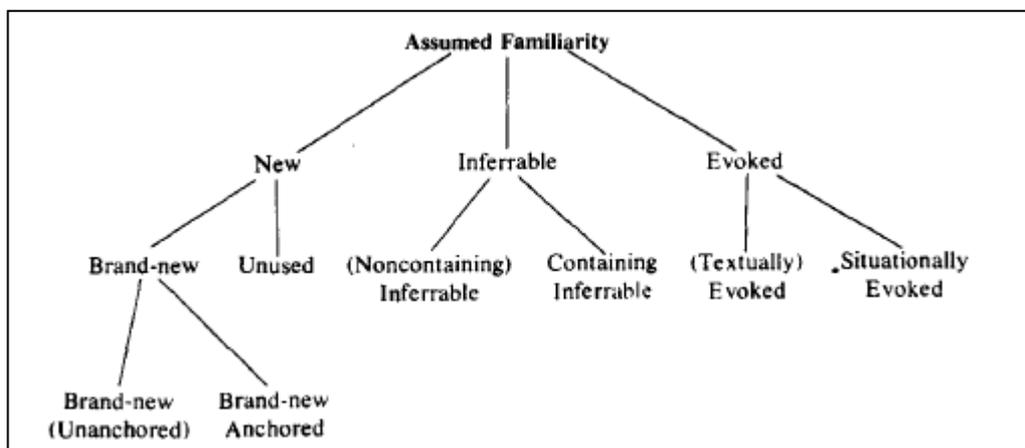
3.2.1 La taxinomie de familiarité supposée de Prince (1981)

Comme évoqué dans la section 3.1 ci-dessus, Prince classe différentes approches de la notion de *givenness* selon leur conception de la notion, en termes de *predictability*, *salience* et *shared knowledge*. L’auteure conclut son examen en proposant que ces trois notions soient intimement liées, et que la conception d’un savoir partagé soit éventuellement nécessaire à la compréhension des deux autres. Toutefois, elle propose de remplacer le terme de savoir partagé par celui de familiarité présumée (*assumed familiarity*), qui aurait l’avantage de se placer du côté des locuteurs, et non comme un observateur extérieur omniscient, puisque « all a speaker has to go on when treating something as given_k [= *givenness* sous l’approche du *shared knowledge*] or "shared" is what s/he assumes the hearer assumes. » (Prince, 1981 : 232). Prince propose alors une taxinomie qui rendrait compte des présuppositions du locuteur ainsi que des inférences auxquelles procédera l’interlocuteur, dans le cadre d’un modèle discursif qui rend compte du statut informationnel des référents en tant qu’entités de ce discours :²⁷

« Turning back to discourse, let us say that a TEXT is a set of instructions from a speaker to a hearer on how to construct a particular DISCOURSE MODEL. The model will contain DISCOURSE ENTITIES, ATTRIBUTES, and LINKS between entities. » (Prince, 1981 : 235)

Le modèle proposé consiste en trois niveaux, subdivisés à leur tour pour aboutir à sept valeurs distinctes de familiarité présumée pour une entité discursive donnée :

Figure II-1 – Types de familiarité présumée (Prince, 1981 : 237)



²⁷ La notion du modèle discursif chez Prince correspond alors précisément aux présuppositions des locuteurs, construites au fur et à mesure, sur les entités qui peuvent être considérés comme ‘ce dont nous parlons’, ainsi que ce qui en est dit. Il s’agit donc d’un modèle partagé entre locuteur et interlocuteur.

Nous allons suivre Prince qui pose un ensemble de cinq énoncés qui permettront d'expliquer les différents niveaux de sa taxinomie :

Exemple II-10 – Exemple d'illustration de niveaux de familiarité présumée (Prince, 1981 : 233)

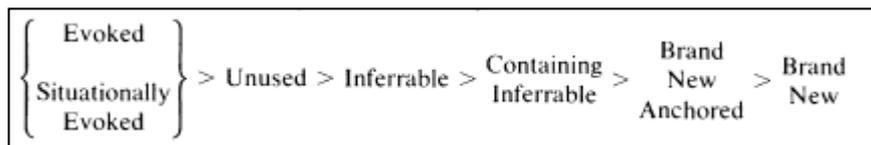
- (a) Pardon, would **you** have change of a quarter?
- (b) **Noam Chomsky** went to Penn .
- (c) I got on **a bus** yesterday and **the driver** was drunk .
- (d) **A guy I work with** says **he** knows your sister .
- (e) Hey, **one of these eggs** is broken!

Prince explique que *New* correspond à l'introduction d'un élément dans le discours. Soit l'interlocuteur devra créer une entité discursive (*Brand-new*), soit il pourra simplement activer une entité déjà connue de lui (*Unused*). Dans l'exemple ci-dessus, *a bus* et *a guy I work with* seront *Brand-new*, tandis que *Noam Chomsky* peut être considéré, en fonction de l'interlocuteur, comme faisant partie de son propre modèle à lui (mémoire à long terme), et correspond donc à *Unused*. Les sous-types de *Brand-new* sont illustrés par *a bus* (*Brand-new unanchored*) et *a guy I work with* (*Brand-new anchored*) respectivement. Dans le deuxième cas, la nouvelle entité, *a guy*, pourra être rattachée au modèle discursif commun par le moyen du lien ou ancrage que constitue le pronom *I*. Cette entité n'est donc pas aussi nouvelle que l'est *a bus*.

Une entité de type *Inferrable* suppose que l'interlocuteur puisse identifier le référent en procédant à des inférences logiques ou plausibles (voir les remarques sur l'anaphore associative et l'exemple de l'église en section 3.1.3 ci-dessus). *The driver* est donc un exemple pour une entité (*Noncontaining*) *Inferrable*, inférable à partir de la mention de *a bus* et du schéma du bus, à savoir que les bus ont des conducteurs. Le cas particulier des *Containing Inferrable* est illustré par l'exemple (d) ci-dessus, où l'inférable et l'élément à partir duquel inférer font partie du même syntagme nominal : ainsi, dans *one of these eggs*, en supposant que *these eggs* soit saillant dans le contexte, *one of these eggs* est inféré de *the eggs*.

Le dernier type de familiarité est *Evoked*, une entité qui est donc déjà supposée se trouver dans le modèle discursif de l'interlocuteur, avec deux sous-types, la mention préalable dans le discours (*(Textually) Evoked*) comme pour le pronom *he* dans l'exemple (d), qui reprend la mention préalable de *a guy I work with*, ou la saillance situationnelle (*Situationnally Evoked*), exemplifié par *you* dans l'exemple (a) et *these eggs* dans l'exemple (e) ci-dessus.

Prince propose ensuite une échelle hiérarchique qui présume qu'un locuteur, pour être coopératif, choisira la forme linguistique la plus adaptée au modèle discursif partagé. Si le locuteur pourrait, en fonction de ce qu'il présume du modèle discursif partagé, utiliser un niveau de familiarité donné, mais choisit de recourir à un niveau inférieur, il enfreint la maxime de quantité de Grice : dire par exemple *J'ai rencontré **une personne*** lorsqu'on peut dire *J'ai rencontré **Marie*** contreviendrait au principe de coopération.

Figure II-2 – Echelle de familiarité (Prince, 1981)²⁸

L'auteure assume et observe une corrélation entre formes linguistiques et l'échelle de familiarité, plus haut une entité se trouve sur l'échelle de familiarité (i.e., le plus à gauche), plus faible pourra être la forme linguistique employée pour y référer, avec des pronoms à un extrême de l'échelle, une description indéfinie simple à l'autre :

Exemple II-11 (Prince, 1981 : 245)

- | | | |
|-----|---|-------------------------|
| (a) | I bought a Toyota. | (Situationally Evoked) |
| (b) | Ellen bought a Toyota. | (Unused) |
| (c) | One of the people that work at Penn bought a Toyota. | (Containing Inferrable) |
| (d) | A person that works at Penn bought a Toyota. | (Brand New Anchored) |
| (e) | A person bought a Toyota. | (Brand New) |

L'auteure observe également une corrélation entre familiarité et fonctions syntaxiques, la fonction sujet étant largement réservé aux entités évoquées dans un discours oral.

« In summation, it seems that we can distinguish a phenomenon that is perhaps true of discourse in general, the tendency to use an NP that is as high on the Familiarity Scale as felicitously possible, and one related phenomenon that seems especially pertinent to informal conversational discourse, a conspiracy of syntactic constructions resulting in the nonoccurrence of NPs low on the scale in subject position. » (Prince, 1981 : 247)

Ces constructions syntaxiques censées éviter qu'une entité qui se situerait plutôt en bas de l'échelle de familiarité se trouve en fonction de sujet syntaxique seraient par exemple « existential *there*, *it*-clefts, Left Dislocation out of subject position, relative clauses, and a variety of so-called "run-on" sentences [...] » (1981 : 246; l'italique est de l'auteure).

Si le modèle de Prince contribue à expliquer des variations dans la construction des énoncés et l'utilisation des expressions référentielles en fonction des besoins communicatifs des interlocuteurs, elle ne recourt pas à la notion de thème ou topic pour le faire. Dans une publication ultérieure (Prince, 1998), elle s'intéressera au lien entre forme et fonction à partir de constructions syntaxiques réputées comme marqueurs de topic. Nous y reviendrons dans la section 4.3.3 ci-dessous.

3.2.2 Les trois états d'activation de Chafe (1987)

Dans son article de 1987, Chafe propose d'affiner son système des statuts informationnels des référents que nous venons d'esquisser dans la section précédente. Bien que la conception de base

²⁸ Pour des raisons de clarté, nous avons choisi la version de l'échelle reproduite dans Gundel et al. (1993 : 280), l'échelle dans Prince (1981) fait figurer seulement des abréviations.

reste inchangée, Chafe formule ces notions désormais en tant que 'états d'activation', et non plus en termes de *consciousness*. Il introduit notamment un niveau intermédiaire par rapport aux états cognitifs de *given* et *new*, et propose les trois états d'activation *active* (*given* dans les travaux antérieurs), *semi-active* (qu'il appelle aussi *accessible*) et *inactive* (*new* dans les travaux antérieurs) :

« There is evidence that a particular concept, at a particular time, may be in any one of three different activation states, which I will call "active", "semi-active", and "inactive". An active concept is one that is currently lit up, a concept in a person's focus of consciousness. A semi-active concept is one that is in a person's peripheral consciousness, a concept of which a person has a background awareness, but which is not being directly focused on. An inactive concept is one that is currently in a person's long-term memory, neither focally nor peripherally active. » (Chafe, 1987 : 25)

Le nouvel état intermédiaire, *semi-active* couvre des mentions d'un référent qui a déjà été mentionné dans le discours antérieur, mais dont les locuteurs n'ont pas parlé depuis un certain moment lorsqu'il est réintroduit. Selon Chafe, les concepts cessent d'être à l'état *active* assez rapidement s'ils ne sont pas utilisés, mais ne passent pas tout de suite à l'état *inactive*. Outre des référents réintroduits, un élément peut avoir le statut de *semi-active* en vertu de son appartenance à un schème ou *script* (Chafe se base ici sur Bartlett (1932; *schema*) ; Mandler & Johnson (1977; *schema*) ; Schank & Abelson (1977; *script*) et Tannen (1979; *schema, script, frame*). Un schème correspond alors à un « cluster of interrelated expectations » (Chafe, 1987 : 29). Le schème du village impliquera des maisons, des rues, une église, une mairie, entre autres ; celui de la voiture impliquera un châssis, des roues, des sièges, volant, phares, et ainsi de suite. Selon Chafe alors, « when a schema has been evoked in a narrative, some if not all of the expectations of which it is constituted presumably enter the semi-active state. » (1987 : 29).

3.2.3 L'échelle d'accessibilité d'Ariel (1988, 1990)

Avec l'échelle d'accessibilité d'Ariel (1988, 1990), l'intérêt se situe clairement non plus, de façon générale, du côté de l'agencement de l'information dans l'énoncé, mais dans la distribution des types d'expressions référentielles. Ariel (1990 : 5) situe son approche comme une réponse à la question suivante : « what conditions allow an appropriate use of all forms marking Accessibility, and moreover, what function is served by accessible information ». Pour Ariel, le concept d'Accessibilité devrait remplacer celui de *givenness*.

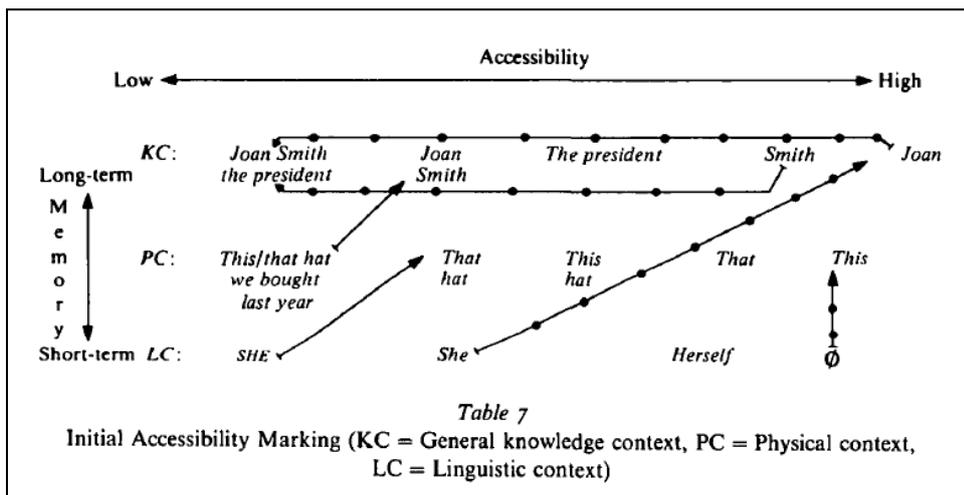
L'auteure propose de combler deux lacunes : d'une part, dépasser le clivage artificiel entre les théories s'intéressant uniquement aux syntagmes nominaux définis et travaillés à partir de la notion de présupposition, et les théories s'intéressant davantage à l'anaphore et spécifiquement aux pronoms (cf. Ariel, 1988 : 65) ; d'autre part, approfondir les travaux existants sur la notion de *givenness*, et explicitement l'échelle de Prince (1981), en associant des formes linguistiques aux

différents niveaux de l'échelle. Selon Ariel (1988), trois types de contexte, à savoir les connaissances du monde (*general knowledge*), le contexte situationnel (*physical surroundings*) et le discours préalable (*previous linguistic material*), correspondent à degrés d'accessibilité du référent. À la base des échelles d'Ariel se trouve donc la question de la récupération (*retrieval*) de la représentation mentale d'un référent :

« Context material is distinguished as to the physical source believed to contain it: General or Encyclopaedic Knowledge, the Physical Environment of the speech event, and the Linguistic Context, i.e. previous utterances actually mentioned in the discourse. Since it is naive to assume that referring expressions directly refer to physical entities (be they linguistic or other kinds of objects), we must assume that in all cases an addressee looks for antecedents which are themselves mental representations. » (Ariel, 1990 : 5-6)

L'échelle d'Accessibilité reproduite ci-dessous représente en conséquence les marqueurs associés à leur contexte d'« initial retrieval ». Cependant, Ariel (1988, 1990) précisera que le concept le plus pertinent pour rendre compte de la distribution des expressions est le degré d'Accessibilité, et non le type de contexte, même si les deux sont liés.

Figure II-3 – L'échelle d'accessibilité (Ariel, 1988 : 81)



Sur cette échelle, les référents antécédents correspondant à des connaissances du monde ou *Knowledge Contexte* (KC) représentent le type de contexte le moins accessible, le contexte linguistique (LC) le plus accessible, le contexte situationnel ou physique (PC) se trouvant à un niveau intermédiaire.²⁹ Pour chaque type de contexte, des types d'expressions référentielles sont classés en fonction de l'accessibilité du référent : du moins accessible (à gauche dans le schéma) au plus accessible (à droite dans le schéma). Au niveau global, un référent qui vient d'être mentionné sera

²⁹ « potential contexts in their natural order of Accessibility (beginning with the previous utterance and proceeding to extend the context by considering utterances further away, assumptions on the current physical situation and Encyclopaedic Knowledge) » (1990 : 167) ; cf. Part III dans Ariel (1990) pour plus de détails sur la notion de contexte.

donc hautement accessible ; si au contraire il faut aller chercher l'antécédent dans les connaissances générales, le référent n'est pas très accessible. Les marqueurs du degré d'accessibilité le plus faible se trouvent donc en haut à gauche dans la figure ci-dessus (*Joan Smith the president*, une description nominale d'un haut degré d'informativité), les marqueurs d'une accessibilité forte en bas à droite (\emptyset , forme zéro). Les flèches simples signalent une double fonctionnalité de certains types d'expressions référentielles. Dans le contexte LC, après mention préalable, le pronom accentué *SHE* marque un degré d'accessibilité plus faible que son équivalent non accentué. Les exemples suivants illustrent cette analyse :

Exemple II-12 – Pronom accentué versus non-accentué en fonction de l'accessibilité en contexte de mention préalable (adapté de Ariel, 1988 : 79)

- (a) Jane_i kissed Mary_j and then she_i kissed Harry.
- (b) Jane_i kissed Mary_j and then SHE_j kissed Harry.

Dans les travaux sur la résolution des pronoms, une préférence pour des antécédents sujet et pour un parallélisme syntaxique (sujet-sujet) a été montrée dans la littérature (pour une synthèse, voir par exemple Colonna, Schimke, & Hemforth, 2012; Kaiser, 2006). Dans l'exemple ci-dessus, *Jane* est donc un antécédent plus facilement accessible pour le pronom de la seconde proposition, ou, autrement dit, le pronom accentué *SHE* en (b) fonctionne comme une instruction pour l'interlocuteur de rejeter l'hypothèse de l'antécédent le plus accessible (i.e. ici le sujet, *Jane*), et de chercher un autre antécédent, ici donc *Mary*. Si la préférence pour le parallélisme syntaxique nous semble valoir aussi pour d'autres langues, cet exemple particulier et le contraste entre pronom accentué et non-accentué vaut bien sûr avant tout pour l'anglais. D'autres langues pourraient utiliser les mêmes ou d'autres procédés pour contrer cette préférence (p.ex. dislocation et clivées en français, l'opposition entre pronoms personnels et démonstratifs en allemand, accentuation : ce point sera développé dans notre CHAPITRE IV).

Le pronom accentué constitue donc, dans le contexte d'un antécédent linguistique mentionné préalablement, un marqueur d'accessibilité relativement faible. Ce même pronom accentué, énoncé dans un contexte où le référent est à chercher dans l'environnement physique, serait un marqueur d'une accessibilité moyenne : ainsi, dans les exemples suivants, énoncés en regardant une personne, et, comme le note Ariel (1988), éventuellement accompagné d'un geste de pointage, *SHE* est un marqueur d'accessibilité plus forte que le serait par exemple une description définie démonstrative comme *cette femme là-bas*.

Exemple II-13 – Pronom accentué comme marqueur d'accessibilité moyenne en contexte physique (adapté de Ariel, 1988 : 80)

- (a) Who's SHE ?
- (b) C'est qui, ELLE ?

Les flèches dotées de points dans la figure ci-dessus indiquent, quant à elles, des emplois possibles et occasionnels de marqueurs de forte accessibilité pour des entités en principe peu accessibles, dépendant selon Ariel très fortement du contexte. Pour comprendre cela, il suffit d'imaginer la

situation d'un parent qui découvre que les chaussures qu'il a achetées à son enfant il y a deux semaines sont trouées, et, les brandissant, s'exclame : *déjà abîmées !* La forme zéro, marqueur de forte accessibilité en contexte linguistique (LC), se trouve ainsi, dans certaines conditions, marqueur de forte accessibilité en contexte physique (PC), au même titre qu'un pronom déictique (nous verrons dans le CHAPITRE IV-2 que ce partage de fonctions est pertinent aussi pour l'emploi des pronoms démonstratifs et formes Zero en allemand).

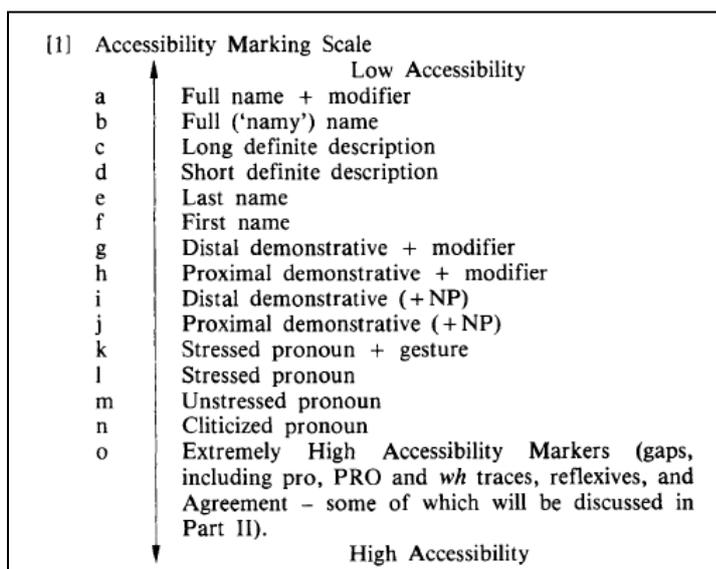
Ariel (1988) propose alors une échelle simplifiée, qui tient compte du fait que le degré d'accessibilité prédit la distribution des expressions référentielles avec plus de précision que les types de contexte :

Figure II-4 – Echelle d'Accessibilité sans référence au type de contexte (Ariel, 1988 : 84)

Joan Smith, the president > Joan Smith > The president > Smith > Joan > That/this hat we bought last year > That hat > This hat > That > This > SHE > she > herself > Ø

Dans sa monographie « *Assessing Noun-phrase Antecedents* », Ariel (1990) propose une échelle plus abstraite que les exemples précédents :

Figure II-5 – Echelle de Marquage de l'Accessibilité (Ariel, 1990 : 73)



Pour Ariel, cette hiérarchie est universelle : l'auteure évoque l'échelle de topicalité de Givón (1983), qui ressemble beaucoup à celle qu'elle propose. L'échelle de Givón représente en effet des résultats d'une étude cross-linguistique. Comme le note Ariel (1988), il n'est pas clair si l'échelle de Givón concerne les expressions référentielles en général, ou uniquement des topics.³⁰ Quoiqu'il en soit, pour préserver la cohérence des sections de ce chapitre, nous ne présenterons pas l'échelle de Givón ici, mais plutôt en section 4 ci-dessous, dédiée à la notion de topic.

³⁰ Tel que nous comprenons les positions de Givón, il nous semble cependant que ses analyses concernent bien tous les référents, l'échelle de topicalité étant une mesure de l'accessibilité des référents, et la notion de topicalité chez Givón n'étant justement pas catégorique, mais scalaire (Givón, 1983, 1988)

L'auteure précise toutefois que son échelle d'accessibilité est spécifique à la langue anglaise : ce qui est universel, c'est le principe qui la sous-tend, notamment la traduction du concept cognitif d'Accessibilité en une corrélation de forme-fonction, par le biais des critères d'informativité, de rigidité et d'atténuation ou poids phonologique.³¹ Elle explique que

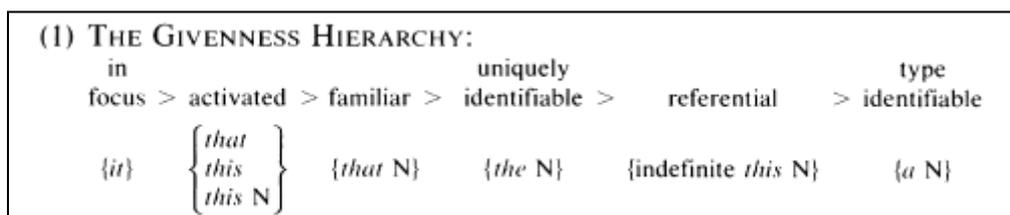
« Basically, the prediction is that although the precise Accessibility rate attached to a specific referring expression may vary from one language to another, no language having counterparts (in the relevant senses, to be indicated below) of the expressions listed in [1= reproduit dans notre Figure II-5 ci-dessus] can arrange them in a different order of Accessibility. It may use them under different circumstances (not pertaining to Accessibility), it may not have some of the expressions listed in [1], it may have others English does not have, but it is predicted not to violate the principle that degree of Accessibility dictates formal choices. » (Ariel, 1990 : 75-76)

3.2.4 Given/new dans les travaux de Gundel et al.

3.2.4.1 La 'givenness hierarchy' de Gundel, Hedberg & Zacharski (1993)

Le modèle de Gundel, Hedberg & Zacharski (1993) partage avec les modèles précités le fait d'expliquer des formes linguistiques, et ici particulièrement le choix des expressions référentielles, en fonction du « assumed cognitive status of the referent, i.e. on assumptions that a cooperative speaker can reasonably make regarding the addressee's knowledge and attention state in the particular context in which the expression is used » (1993 : 275). La *Givenness Hierarchy* est alors une échelle de statuts cognitifs présumés, définis comme « information about location in memory and attention state » (1993 : 274). Nous reproduisons leur exemple (1) en Figure II-6 ci-dessous :

Figure II-6 – La Givenness Hierarchie de Gundel, Hedberg & Zacharski (1993 : 275)



Les auteurs précisent qu'à la différence d'autres modèles comparables, les statuts ne sont pas mutuellement exclusifs, mais un statut donné implique aussi tous les statuts inférieurs (alors que l'inverse n'est pas vrai) : si le locuteur peut supposer que la représentation mentale d'un référent a, par exemple, le statut *activated* pour l'interlocuteur, cela vaudra également pour les statuts inférieurs (à droite dans le schéma). Un référent activé sera aussi familier, identifiable de façon unique, et ainsi de suite. Nous proposons un rapide résumé des différents statuts :

³¹ L'auteure admet des exceptions à cette corrélation : « Only marginal instances were found where the form-function correlations were arbitrary, i.e. due to a conventional association of a specific marker with a certain degree of Accessibility. » (1990 : 92)

Type identifiable signifie que l'interlocuteur peut accéder à une représentation mentale d'une classe de référents à partir de la signification lexicale d'une description nominale : les auteurs expliquent qu'une expression comme *a dog* est appropriée lorsque le locuteur connaît le sens de ce mot. Le statut *referential* s'applique selon les auteurs au *colloquial English*, et il faut le comparer au statut *type identifiable* : A la différence de la description indéfinie *a dog* ou *un chien* en français, qui permettrait soit une interprétation référentielle (le locuteur affirme quelque chose d'un chien particulier), soit une interprétation non-référentielle (le locuteur affirme seulement l'existence d'au moins un exemplaire de la classe *chien*), lors de l'emploi d'une expression de type *referential* (indefinite *this N*), seule une interprétation référentielle serait possible. Gundel et al. mentionnent des résultats de recherches montrant que des descriptions nominales indéfinies avec le déterminant *this* (à interprétation indéfinie, voir les exemples ci-dessous) seraient plus susceptibles d'être mentionnés dans la suite d'un discours. Les mêmes réflexions peuvent être faites d'ailleurs pour le français et pour l'allemand :

Exemple II-14 – Descriptions indéfinies avec article indéfini et déterminant démonstratif en français et en allemand

- | | | |
|-----|---|--|
| (a) | Hier, un type m'a fait un sourire dans le métro. | Gestern hat mich ein Typ in der Metro angelächelt. |
| | (a1) Il avait l'air sympa. | Der sah nett aus. |
| | (a2) Aujourd'hui, personne. | Heute keiner. |
| (b) | Hier, ce type m'a fait un sourire dans le métro. | Gestern hat mich dieser Typ in der Metro angelächelt. |
| | (b1) Il avait l'air sympa. | Der sah nett aus. |
| | (b2) ?? Aujourd'hui, personne. | ?? Heute keiner. |

Si les énoncés en (a) peuvent être compris comme parlant d'un type en particulier (*cf.* la suite possible en (a1)), soit juste affirmant l'existence d'une telle personne (*cf.* la suite possible en (a2)), les énoncés en (b) peuvent seulement être interprétés comme faisant référence à une personne particulière (et la suite en (b2) devient difficilement acceptable), bien que l'interlocuteur ne soit pas plus en mesure d'identifier le référent particulier qu'il ne l'est avec les énoncés en (a). Ce n'est qu'à partir du statut suivant, *uniquely identifiable*, condition nécessaire pour les descriptions définies, que l'interlocuteur pourra identifier le référent particulier dont parle le locuteur, sur la base de la seule mention de l'expression référentielle. Ainsi, les auteurs expliquent que *the dog next door* est une instruction suffisante pour que le locuteur crée une représentation mentale du chien des voisins, même s'il ignorait que les voisins en possédaient un. Au contraire, pour qu'un référent puisse avoir le statut *familiar*, l'interlocuteur doit, comme le dit le nom, avoir une représentation préalable du référent en question : *that dog next door* ne pourrait, selon les auteurs, être interprété par l'interlocuteur que s'il a déjà connaissance de ce chien. Un référent qui se trouve dans la mémoire à court terme (parce qu'il aura été mentionné récemment ou parce qu'il fait partie de l'environnement physique immédiat) aura le statut de *activated*, tandis qu'un référent du statut *in focus* se trouvera dans le « current center of attention » (1993 : 279).

Gundel et al. (1993) comparent leur hiérarchie avec l'échelle de familiarité de Prince (1981). Ils établissent certaines correspondances entre les deux modèles, que nous présentons par le tableau ci-dessous :

Tableau II-1 – Comparaison de la Familiarity Scale de Prince et la Givenness Hierarchy de Gundel et al.

Assumed Familiarity Scale (Prince, 1981)	Givenness Hierarchy (J. K. Gundel et al., 1993)
Evoked + Situationally Evoked	In focus
	Activated
Unused	Familiar
Inferrable	?
Containing Inferrable	Uniquely identifiable
Brand New Anchored	(Non discuté par les auteurs)
Brand New	Type identifiable

Gundel et al. (1993) proposent que la catégorie Inferrable de Prince n'est pas assez précise et pourrait correspondre à différents statuts dans la givenness hierarchy, allant du *type identifiable* au *familiar*, en fonction de la force du lien d'inférence : « We would thus expect inferrables to have different statuses, and to be coded by different forms, depending on the nature and strength of the link between the inferrable and its associated discourse entity » (J. K. Gundel et al., 1993 : 281).

En ce qui concerne le lien entre les statuts cognitifs et les formes linguistiques, chaque statut constitue donc des conditions suffisantes et nécessaires pour certaines expressions référentielles, qui ne pourront être employées de façon appropriée à un niveau inférieur (un pronom ne sera pas approprié si le référent a seulement le statut de *type identifiable*, par exemple). En revanche, une expression référentielle appropriée pour un niveau donné pourra également référer à tous les niveaux supérieurs (pour l'emploi d'une description définie, le référent devra être au moins *uniquely identifiable*, mais elle pourra également être employée pour des statuts plus élevés, pour une référence au statut *activated*, par exemple). Les auteurs se servent de la maxime de quantité de Grice (1975) pour expliquer la juste quantité d'information que le locuteur devra fournir à l'interlocuteur.

Enfin, comme Prince (1981) et Ariel (1988, 1990), Gundel et al. discutent la validité universelle de leur modèle, en comparant des résultats d'analyses effectuées pour l'anglais, le chinois, le japonais, le russe et l'espagnol (données écrites et orales, à des degrés de formalité diverses, cf. Gundel et al. (1993 : 290-291)).

3.2.4.2 La distinction de deux types de givenness dans les travaux de Gundel et Lambrecht

Pour clore cette section, nous évoquerons assez rapidement les clarifications apportées par Gundel (1988, 1999), défendues de façon similaire dans Lambrecht (1994) concernant la notion de *givenness*. Selon les auteurs, il est nécessaire de distinguer entre *referential givenness* et *relational givenness*. Cette distinction est partiellement appliquée dans les réflexions de Chafe, Halliday, Kuno (cf. ci-dessus), mais Gundel et Lambrecht la relieront explicitement aux difficultés d'aboutir à une définition satisfaisante de thème/topic, parce que ces deux types de *givenness* seraient justement souvent confondus :

« Much of the confusion surrounding these issues has resulted from conflating two types of givenness/newness [...]. One type is referential; it involves a relation between a linguistic expression and a corresponding non-linguistic entity in the speaker/hearer's mind, the discourse, or some real or possible world, depending on where the referents or corresponding meanings of these linguistic expressions are assumed to reside. [...]

The second type of givenness/newness is relational. It involves two complementary parts, X and Y, of a linguistic or conceptual representation, where X is given in relation to Y, and Y is new in relation to X. » (J. K. Gundel, 1999 : 188)

L'auteure précise que la première acception correspond aux statuts cognitifs de référents, tandis que la seconde, indépendante de la première, appartient à la dimension topic/comment, qui nous occupera dans les sections suivantes.

3.3 Synthèse

Nous avons dressé dans cette section un tableau partiel des différentes approches qui se sont consacrées à la structuration de l'information, et notamment à la distinction entre information donnée/ nouvelle ainsi que son expression linguistique. Dans le déroulement chronologique des travaux, l'on a pu noter une complexification des modèles, les modèles binaires du type *old/new* faisant place à des modèles ou échelles d'une plus grande finesse de degrés. Dès les premiers travaux, l'on peut observer une plus ou moins grande précaution face à la notion d'information donnée et ses relations avec d'autres niveaux de la description linguistiques, comme la fonction de sujet grammatical, l'ordre des mots dans l'énoncé ou la notion, cernée de façon plutôt vague, de thème/topic : Ainsi, si certains auteurs ont pu proposer une coïncidence définitoire entre ces différents domaines le consensus semble aller à la reconnaissance d'une tendance qui laisse place à des divergences : sujet, information donnée et topic coïncidant souvent, mais non pas nécessairement. Avec d'autres travaux, exemplifiés ici par ceux d'Ariel (1988, 1990) et Gundel et al. (1993), le centre d'intérêt s'est déporté des questions de l'ordre de mots et structuration de l'information vers la question des expressions référentielles et leur façon de rendre des statuts cognitifs ou informationnels (au sens large ici, adapté selon chaque approche particulière). À travers les positions de Halliday, Chafe et Kuno notamment, l'on peut déjà pressentir comment les dimensions de l'information donnée et celle de thème/topic s'entrelacent. À quel point et de quelles manières, c'est ce que nous tenterons d'exposer dans la section suivante, qui sera dédiée à la notion de thème/topic.

4. La notion de thème/topic

Topic, topique, thème : de quoi parle-t-on ? Les termes sont multiples, les notions qu'ils traduisent le semblent encore davantage. Parfois présentés comme des synonymes ou termes équivalents, parfois

utilisés comme concepts complémentaires, ce qui ressort d'une revue des travaux est l'hétérogénéité des définitions.

Dans le domaine des travaux portant sur la structuration informationnelle des énoncés et des discours, la notion de topic/topique/thème a été abondamment discutée. Les considérations de cette notion diffèrent pourtant largement : si pour certains thème/topic est une catégorie grammaticale de la phrase au même titre que l'est par exemple la catégorie sujet (p.ex. Li & Thompson, 1976), pour d'autres il s'agit d'une notion à caractère informationnel et interactionnel, qui peut être saisie uniquement au niveau du discours et ne relève pas de la grammaire (cf. par exemple la distinction entre thème et topic chez Berthoud & Mondada (1993)). Afin de nous positionner clairement par rapport aux différentes acceptions terminologiques existantes, nous allons d'abord dresser un tableau des termes et principes définitoires que l'on trouve dans la littérature. Après avoir retracé ensuite rapidement le chemin d'apparition de la notion de thème, nous nous consacrerons à l'examen détaillé des différents types de définition rencontrés.

4.1 La profusion terminologique : un premier constat

Si la notion de topic/thème a suscité un grand nombre de travaux, elle n'en reste pas moins problématique. D'après Brown & Yule (1988 [1983] : 70), « 'topic' could be described as the most frequently used, unexplained, term in the analysis of discourse ». Reinhart (1981) tient des propos similaires :

« Although the linguistic role of the relation *topic of* is widely acknowledged, there is no accepted definition for it, and not even full agreement on the intuitions of what counts as topic. In fact, almost every imaginable approach to its definition has been actually proposed. » (1981 : 56)

D'autres auteurs se joignent à ce constat, et certains vont même jusqu'à proposer l'abandon de la notion de topic (Berrendonner & Reichler-Béguelin, 1997; Levinson, 1983; Schlobinski & Schütze-Coburn, 1992).

La notion de 'topic/thème' est, à l'origine, la première partie d'un couple de notions, traduisant une bipartition de la phrase, comme nous l'avons développé en section 1 ci-dessus. Galmiche (1992) fournit la liste suivante, qui témoigne de la profusion des termes/couplages recensés :

Tableau II-2 – Couplages thème/rhème recensés dans la littérature (adapté de Galmiche, 1992 : 3)³²

Terme 1	Terme 2	Notamment dans :
thème	propos	(Bally, 1944 [1932])
thème	prédicat	---
thème	rhème	(Halliday, 1967; Vallduví, 1993)
topic	comment	(Gundel, 1985, 1974/1988, 1999; Gundel, Hedberg, & Zacharski, 1997; Hockett, 1958; Hornby, 1971, 1972; Lambrecht, 1994)
topique	commentaire	(Cornish, 2008; Horváth, 2018)
topique	propos	(Grobet, 2002)
topic	focus	(Dik, 1997 [1989]; Erteschik-Shir, 1997, 2007)
présupposition	focus	(Chomsky, 1976 [1970]; Jackendoff, 1972)
sujet psychologique	prédicat psychologique	(Gabelentz, 1869, 1891; Paul, 1920 [1880])

Il a été constaté par différents auteurs (cf. i.a. Grobet, 2002 : 17-18 pour une synthèse) que d'une part, la notion de 'thème' n'est pas nécessairement la même dans les différents couplages employés par des approches différentes, et que d'autre part le lien entre les termes thème et *topic* est loin d'être clair, comme le montre Prévost (1998) : les deux peuvent être considérés comme équivalents, *topic* étant l'équivalent anglophone de 'thème' (bien que l'on trouve aussi le terme *topic* (ou *topique*) dans la littérature francophone, comme le rappelle Prévost). Si dans la plupart des approches un seul des termes est utilisé, dans l'approche de Dik (1997 [1989]) les deux sont complémentaires. C'est le couple *topic/focus* qui décrit l'agencement de la proposition, *topic* désignant un constituant régi par le verbe, alors que *theme* et *tail* s'appliquent aux constituants extra-propositionnels et correspondent entre autres aux dislocations (*intra-clausal* vs. *extra-clausal constituent* (cf. Dik, 1997 [1989] : Chap. 13). Plus précisément, *theme* est utilisé pour les extrapositions en début de phrase, *tail* pour les extrapositions à la fin, correspondant aux dislocations gauche et droite.

En lien étroit avec ces problèmes terminologiques, Galmiche (1992) rappelle que c'est au niveau de la description (et il précise qu'il entend par là : « caractérisation, exemplification, tentative de définition » (1992 : 4)) que l'hétérogénéité des notions apparaît réellement, le même terme pouvant être défini à partir d'angles très différents. Ceci, selon lui, est dû à la nature complexe et transversale du phénomène à décrire, touchant à plusieurs niveaux linguistiques : sémantique, logique, psychologique, pragmatique, énonciatif, prosodique, syntaxique. En effet, le second tableau dressé par Galmiche résume bien l'étendue des niveaux auxquels l'on a tenté de décrire (ou définir) le 'thème' :

³² Nous avons ajouté un couple terminologique supplémentaire (en gris dans le tableau), et nous avons ajouté aussi des références qui nous semblaient importantes (le tableau de Galmiche n'en comporte pas), sans prétention d'exhaustivité.

Tableau II-3 – tentatives de description de thème/topic (adapté de Galmiche, 1992 : 4)

Terme 1	Terme 2
ce dont on parle	ce qu'on en dit
ancien	nouveau
donné	non-donné
connu	non-connu
présupposé	focalisé
	emphatique
point de départ	but
base	aboutissement
statique	dynamique
moins informatif	plus informatif
support	apport
fond	figure
récupérable	non-récupérable
prévisible	non-prévisible
activé	
saillant	
centre d'intérêt	
lié à la conscience immédiate	
lié au contexte	indépendant du contexte
lié aux circonstances notoire	

Plutôt de commenter ici chaque entrée de ce tableau, nous allons continuer à dresser le tableau des problèmes méthodologiques, et proposer une discussion synthétique de différents types de définition en section 4.3.

Galmiche soulève un problème supplémentaire, qui vient selon lui du fait que l'on a recours au langage ordinaire pour décrire le phénomène, et il rappelle que chez certains auteurs des termes qui pourront appartenir à la colonne droite sont employés pour décrire la colonne gauche (donc le thème). Ainsi, il rappelle que Reinhart (1981) déjà

« dénonce les descriptions présentées par Schachter (1973) et Garcia (1975) qui, à propos du thème (ou Topic, i.e. versant gauche), parlent d'expressions représentant « le centre ou le Focus d'attention du locuteur ». De même, plus récemment, van Oosten (1986) décrit également le « topic » comme le Focus d'attention » (Galmiche, 1992 : 4)

Et en effet, Reinhart (1981) note que ces mêmes notions psychologiques de 'centre d'attention' ou 'l'intention' du locuteur peuvent aussi bien servir à décrire le complément de l'expression du topic : « the same psychological notions describe just as successfully the complement of the topic expressions - that part of the sentence which is 'dominant' or provides new information about the topic. » (1981 : 57)

De même, Gundel et al. (1993) ont recours à la notion *in focus* pour caractériser des expressions topicales. Ils rappellent qu'il existe effectivement deux acceptions différentes de la notion de *focus*, l'une psychologique, délimitant ce qui se trouve dans le centre d'attention des interlocuteurs (et pouvant être rattaché à la colonne gauche du Tableau II-3 de Galmiche ci-dessus, au *topic*), l'autre

plus formelle (se situant par sa définition dans la colonne droite de ce même tableau): « the position of linguistic prominence in the part of the sentence that expresses the comment » (1993 : 279).

Une clarification terminologique préalable est donc nécessaire. Nous n'avons pas la prétention de proposer ici une solution universelle au problème de la définition de thème/topic, mais proposons une présentation et discussion des définitions les plus fréquemment employées. Nous argumenterons pour celle qui nous paraît comme la plus opérationnelle pour la tâche que nous nous proposons dans le présent travail de thèse. Dans ce qui suit, nous allons en conséquence, et pour circonscrire le domaine à étudier, nous intéresser prioritairement à la première partie des couples notionnels présentés, au thème/topic.

4.2 Une petite histoire du 'thème'

Avant de discuter dans le détail les différents types de définitions qui ont été proposés, nous souhaitons revenir brièvement aux origines de la notion de thème/topic. Nous l'avons déjà esquissé en section 1, cette notion est l'héritage de la logique antique. Le couple thème/rhème, décrit à partir de la phrase canonique et isolée, repose sur la mise en correspondance des dimensions logique et syntaxique : les choses et ce que l'on en dit, le nom et le verbe, le sujet et le prédicat (cf. Siblot, 1998). Le fait que thème et sujet grammatical ne coïncident pas nécessairement est introduit dans les sciences du langage avec la notion de 'sujet psychologique' (cf. section 1). Dans la littérature française, c'est à Bally (1944 [1932]) que l'on attribue l'introduction des notions 'thème' et 'propos', avec référence explicite aux 'sujet et prédicat psychologiques' :

« la pensée qu'on veut faire connaître est – nous le répétons, cf. 32 – le but, la fin de l'énoncé, ce qu'on se propose, en un mot : le *propos* ; on l'énonce à l'occasion d'une autre chose qui en forme la base, le substrat, le motif : c'est le *thème*. »
(Bally, 1944 [1932] : 53)

À la même période, aux débuts du Cercle Linguistique de Prague, Mathesius introduit le concept de la *Functional Sentence Perspective* (1929). Dans ce cadre théorique, il développe le couple *Satzthema/Mitteilungskern* ('*Basis of the statement/Nucleus of the statement*') (Mathesius, 1971 [1929]). Dans les écrits du Cercle de Prague, on trouve également les termes '*theme/rheme*' pour ces notions : « The element about which something is stated may be said to be the basis of the utterance or the theme, and what is stated about the basis is the nucleus of the utterance or the rheme. » (Mathesius, 1975 [1961] : 81).

Le terme *topic*, quant à lui, est souvent considéré comme la traduction, dans les travaux anglophones, du terme thème. Son apparition est souvent attribuée à Hockett (1958), qui définit la notion en termes de *aboutness* : « The most general characterization of predicative constructions is suggested by the terms "topic" and "comment" for their ICs [=immediate constituents]: the speaker

announces a topic and then says something about it » (1958 : 201). L'auteur remarque également que dans des langues comme l'anglais, l'organisation de *topic/comment* coïncide préférentiellement avec celle de sujet/prédicat, alors que cette association préférentielle ne tiendrait pas pour le chinois, par exemple (1958 : 202).

Il nous semble important de noter que dès les premières analyses de la dimension thème/propos, des critères de différents ordres interviennent pour définir le thème : ainsi, dès les premiers travaux, l'on trouve une définition en termes de 'à propos' ou *aboutness*. Nous rappelons que Gabelentz (1891 : 353-354) dit ainsi « dass ich erst dasjenige nenne, was mein Denken anregt, worüber ich nachdenke, mein psychologisches Subject, und dann das, was ich darüber denke, mein psychologisches Prädicat [...] » (1891 : 353-354).³³ De la même manière se trouvent déjà chez Paul les conceptions du prédicat psychologique (ou rhème) comme nouveau (« *das neu hinzutretende* » ou comme plus important (« *das bedeutsamere* » (1920 [1880] : 126). Firbas (1957) de même, en se basant sur les travaux de Weil (1844) et Mathesius, met les notions thème/rhème en parallèle avec le couple connu/nouveau :

« Viewed thus, those sentence elements which convey something already known or something that may be taken for granted, in other words, those elements that may be inferred either from the verbal or from the situational context, are to be regarded as the communicative basis of the sentence. They are referred to in this paper as the theme of the sentence [...] On the other hand, those sentence elements which convey the new piece of information are to be regarded as the communicative nucleus of the sentence. They are referred to in this paper as the rheme of the sentence [...] » (Firbas, 1957 : 72)

Pour Firbas, il découle de cette distinction que les éléments thématiques seraient « less important in the given situation, being communicatively less dynamic than the rhematic elements » (1957 : 72).

Il ressort également de ces premiers travaux l'idée que l'agencement de la phrase/ de l'énoncé se reflète dans l'ordre linéaire des mots, tantôt considéré comme principe universel, où le thème comme élément connu précède le rhème comme élément nouveau (cf. Gabelentz (1891); Weil (1844)), tantôt nuancé en fonction des règles syntaxiques de chaque langue (cf. i.a. Firbas, 1957; Paul, 1920 [1880]).

Ces quelques exemples montrent donc que dès le début des réflexions sur la notion de 'thème', les différents principes définitoires mentionnés plus haut sont déjà présents. Différentes approches ont mis l'accent sur l'un ou l'autre pour cerner le thème/topic, et nous allons dans ce qui suit nous consacrer à l'examen de chacun de ces principes de définition dans des approches plus récentes.

³³ « que je nomme d'abord ce qui motive ma pensée, ce à quoi je réfléchis, mon sujet psychologique, et ensuite ce que j'en pense, mon prédicat psychologique [...] » (Gabelentz, 1891 : 353-354; notre traduction)

4.3 Examen des différents principes définitoires du thème/topic

Nous proposons ici une classification des approches selon la logique sous-jacente de la définition de topic : des définitions sur critères formels, qui se basent sur une position dans l'ordre linéaire des mots ou une position syntaxique ; des définitions sur critères informationnels et des définitions sur critères ou pragmatiques ou fonctionnels.³⁴ Étant donné qu'il s'agit globalement d'un même phénomène, le thème/topic, défini à partir de points de vue différents, il n'est pas étonnant de trouver des passerelles entre différents types de définition que nous signalerons au lecteur.

4.3.1 Une définition formelle du topic

Nous allons regrouper dans cette section des définitions de thème/topic qui sont basés principalement sur des critères formels : une position dans l'ordre linéaire des mots (travaux de l'École de Prague, cf. synthèse de Firbas, 1964; Halliday, 1967), ou une position syntaxique (Chomsky, 1965, 1976 [1970]; Jackendoff, 1972; Rizzi, 1997).

Comme nous l'avons déjà montré dans ce chapitre, l'idée que structure informationnelle et organisation syntaxique du discours sont liées se retrouve très tôt dans les réflexions des linguistes qui s'intéressent au phénomène du thème/topic. À titre d'exemple, nous reprenons la citation de Gabelentz (1891), déjà discutée en section 4.2 : « que je nomme d'abord ce qui motive ma pensée, ce à quoi je réfléchis, mon sujet psychologique, et ensuite ce que j'en pense, mon prédicat psychologique [...] » (1891 : 353-354; notre traduction). L'organisation de l'information est donc mise en relation avec l'ordre linéaire des mots. De ces considérations sont nées des définitions du thème/topic sur critères formels, dont la plus souvent citée semble être celle de Halliday (1967).

4.3.1.1 Halliday

Comme le note Grobet (2002), l'approche de Halliday est composée de deux facettes, l'une étant un principe fonctionnel, l'autre sa manifestation formelle :

« The theme is what is being talked about, the point of departure for the clause as a message; and the speaker has within certain limits the option of selecting any element in the clause as thematic. » (Halliday, 1967 : 212)

Le thème est donc défini, au départ, en termes de 'à propos' ou 'point de départ'.³⁵ Quant à l'organisation formelle de ce principe, Halliday note que

³⁴ Cette classification est largement inspirée des synthèses proposées par Gómez González (1997), Grobet (2002) et Schlobinski & Schütze-Coburn (1992).

³⁵ Cette idée est déjà avancée dans la thèse de Weil (1844 : 25).

« [...] thematization takes a unit of sentence structure, the clause, and structures it in a way that is independent of what has gone before. This structuring is into two parts, a theme and a rheme, and is realized simply by the sequence of elements: the theme is assigned initial position in the clause, and all that follows is the rheme. Thus in *John saw the play yesterday*, *yesterday John saw the play* and *the play John saw yesterday* (as a complete clause) the themes are, respectively, *John*, *yesterday* and *the play*. » (Halliday, 1967 : 212)

Le thème est donc, globalement, exprimé par la position initiale dans la proposition (*clause* chez Halliday), qu'il s'agisse d'un syntagme nominal (*John saw the play yesterday*), d'un adverbe (*yesterday John saw the play*), d'un verbe modal (*did John see the play yesterday?*) ou d'un pronom interrogatif (*what did John see yesterday?*).³⁶

4.3.1.2 Chafe

Chafe (1976) fait la distinction entre ce qu'il appelle 'subject' et 'topic' (cf. aussi les commentaires de Lambrecht, 1994 : 118 à ce sujet). 'Subject' est alors défini dans les termes suivants : « identifying some particular as a starting point and adding to the addressee's knowledge about it » (1976 : 44). Chafe emploie également la métaphore du *hitching-post* auquel s'attachera l'information nouvelle. De telles métaphores sont employées également par d'autres auteurs, qui définissent le topic en termes d'*aboutness* (cf. section 4.3.3 ci-dessous). De fait, plus loin, Chafe convient explicitement que ce qu'il appelle 'subject' pourrait correspondre à ce que d'autres auteurs appellent 'topic' (défini comme étant « what the speaker is talking about » (1976 : 47).³⁷ Chafe réserve alors le terme de 'topic' à des constructions spécialisées comme la topicalisation en anglais et les 'chinese style topics'.³⁸ Reprenons, pour illustrer ces derniers, un exemple que Chafe emprunte à Li & Thompson (1976) :

Exemple II-15 – Chinese style topic (Chafe, 1976 : 50)

Nèi-xie shùmu shù-shēn dà
Those tree tree-trunk big

Il faut noter que Chafe attribue des fonctions différentes à ce qu'il appelle *subject* et *topic* en anglais, d'une part, et les *chinese style topics* d'autre part : ces derniers « do not fit precisely the characterization that a topic is "what the sentence is about" », mais leur fonction serait plutôt « to limit the applicability of the main predication to a certain restricted domain » (Chafe, 1976 : 50). L'auteur précise alors cette fonction en tant que cadre : « the [chinese style ; *notre ajout*] topic sets a spatial, temporal, or individual framework within which the main predication holds » (Chafe, 1976 :

³⁶ Exemples tirés de Halliday (1967 : 222). Nous nous contentons ici de résumer les thèmes propres et nécessaires à la proposition, que Halliday résume sous le terme *cognitive*, et ne traitons pas les thèmes supplémentaires possibles (*non-cognitive*), qui relèvent des niveaux de l'énonciation ou de l'organisation discursive (cf. tableau synthétique de Halliday (1967 : 222)).

³⁷ Chafe cite cette définition de Hornby (1971 : 1976).

³⁸ Cette position rappelle d'ailleurs la différenciation que fait Dik (1997 [1989]) entre *topic* et *theme/tail* (cf. section 4.1 ci-dessus); avec la complication terminologique supplémentaire que les *subject* et *topic* de Chafe (1976) correspondraient peu ou prou à *topic* et *theme/tail* respectivement chez Dik.

50). Dans l'exemple ci-dessus, l'énoncé ne serait donc pas une prédication à propos de *those trees*, mais à propos de *tree-trunk*. Nous reviendrons à la distinction entre topic d'à propos et topic cadratif en section 4.3.3.5 ci-dessous.

À travers les exemples donnés dans l'article de Chafe (1976), l'auteur semble indiquer qu'il identifie *subject* et/ou *topic* avec une position linéaire initiale ou finale. Dans une visée typologique, il observe ensuite que dans des langues comme l'anglais, cette relation se traduira typiquement par une fonction syntaxique (celle de sujet grammatical), alors que dans d'autres langues, l'ordre des mots semblerait marquer cette relation.³⁹

Dans une publication plus récente, Chafe (1987) formule ces conceptions à travers le prisme des états d'activation informationnels (cf. section 3.2.2 ci-dessus). Il utilise explicitement les termes « starting point » et « added information » pour ce que d'autres appellent 'topic' et 'comment' (1987 : 22). Dans sa définition de ces niveaux d'organisation de l'information, il les lie à la structuration grammaticale 'sujet-prédicat' :

« A speaker does not simply thrust concepts forward out of nowhere. The usual technique for presenting information is to choose some concept, typically some referent, as a starting point and then to add information about it. As a speaker proceeds to verbalize one focus of consciousness after another, each added piece of information is attached to some other piece that is in some sense already present. The linguistic manifestation of this formatting strategy is the familiar subject-predicate structure. Many information units conform to this structure, employing a subject to express the starting point and proceeding with a predicate which adds information about that starting point. » (Chafe, 1987 : 36)

Concernant la composante qui nous intéresse ici, le 'starting point' ou le 'topic', il le met en relation avec les trois états d'activation présentés en section 3.2.2 ci-dessus sous forme d'une contrainte hypothétique, qui stipule qu'un 'starting point' correspondra typiquement à un référent donné, occasionnellement à un référent accessible et rarement à un référent nouveau (cf. 1987 : 37).

Ces réflexions sont étayées par l'analyse d'un extrait d'une conversation informelle entre adultes. Cependant, nous notons qu'il s'agit d'une anecdote racontée seulement par l'un des participants, qui constitue ainsi un petit monologue de taille réduite (environ 10 énoncés ou 40 unités intonatives). D'une part, il paraît très difficile de tirer une quelconque généralisation d'un tel corpus, qui ne peut avoir autre fonction que celle d'exemplifier, d'autre part, les observations de Chafe sur le 'topic' ou 'starting point' constituent une description de certains aspects des topics dans un monologue donné, mais en aucun cas une définition.

³⁹ La définition même que donne Chafe à la notion de topic (*subject* et *topic* chez lui) pourra donc également être classée parmi les approches pragmatiques de la section 4.3.3. Nous le classons cependant plutôt parmi les approches formelles, puisque les corrélations qu'il établit entre subject/topic et des caractéristiques formelles des langues telles que l'ordre des mots et les fonctions syntaxiques sont formulées de manière plutôt forte.

4.3.1.3 Dans les travaux générativistes

Dans les travaux générativistes, le 'topic' correspond également à une position syntaxique. Chomsky (1965) traite cette dimension seulement dans une note marginale, et l'associe au premier élément dans la structure de surface d'une phrase, qui coïncidera souvent, mais non nécessairement avec le sujet grammatical de cette phrase :

« It might be suggested that Topic-Comment is the basic grammatical relation of surface structure corresponding (roughly) to the fundamental Subject-Predicate relation of deep structure. Thus we might define the Topic-of the Sentence as the leftmost NP immediately dominated by S in the surface structure, and the Comment-of the Sentence as the rest of the string. Often, of course, Topic and Subject will coincide, but not in the examples discussed. » (1965 : 221)

Les termes *Topic/Comment* feront place, dans des écrits ultérieurs, à l'articulation *presupposition/focus* (Chomsky, 1976 [1970]). Pour Hajičová & Sgall (1975) *presupposition* chez Chomsky (1976 [1970]) ainsi que chez Jackendoff (1972) est globalement l'équivalent de *topic* :

« We suppose that Chomsky's dichotomy of presupposition (or topic) and focus corresponds (with reservations [...]) to the traditional dichotomy of psychological (or logical) subject and predicate, or, in more modern terms, of theme and rheme, topic and comment. » (Hajičová & Sgall, 1975 : 7-8)

Il nous semble cependant que, indépendamment des similarités dans l'intension, l'extension de *presupposition/focus* ne correspond pas à celle de sujet/prédicat psychologique tel qu'envisagée par Gabelentz (1869, 1891) et Paul (1920 [1880]). En effet, Chomsky donne la définition suivante :

« The focus is a phrase containing the intonation center; the presupposition, an expression derived by replacing the focus by a variable. » (Chomsky, 1976 [1970] : 100)

L'auteur explique que l'étendue du focus se détermine par ce qui constitue une « réponse naturelle » (1976 [1970] : 93; notre traduction), et que la *presupposition* « can be determined by replacing what is taken as focus by an appropriate variable » (1976 [1970] : 94) comme illustré dans l'exemple suivant :

Exemple II-16 – tiré de Gundel (1999 : 192)⁴⁰

Did the Red Sox play the YANKEES?

PRESUPPOSITION

the RS played someone
the RS did something
something happened

FOCUS

the Yankees
played the Yankees
the RS played the Yankees

POSSIBLE RESPONSE

No. The Tigers.
No. They had the day off.
No. Bill had the flu.

Il est également intéressant de noter que Jackendoff (1972), qui développe une composante sémantique dans la grammaire transformationnelle, définit *presupposition/focus* davantage en termes de savoir partagé :

« As working definitions, we will use “focus of a sentence” to denote the information in the sentence that is assumed by the speaker not to be shared by him and the hearer, and “presupposition of a sentence” to denote the information in the sentence that is assumed by the speaker to be shared by him and the hearer. » (Jackendoff, 1972 : 230)

Rizzi (1997) s'intéresse à la périphérie gauche de la phrase/ de l'énoncé dans différentes langues, et se limite à l'examen de structures spécialisées pour le marquage de la structure informationnelle, comme la topicalisation et la dislocation. Il propose une définition de 'topic' qui semble essentiellement syntaxique :

« The topic is a preposed element characteristically set off from the rest of the clause by “comma intonation” and normally expressing old information, somehow available and salient in previous discourse; the comment is a kind of complex predicate, an open sentence predicated of the topic and introducing new information. » (Rizzi, 1997 : 285)

Même si des notions d'ordre informationnel et discursif entrent dans cette définition du topic, la condition essentielle semble être l'antéposition en surface en ce qui concerne la périphérie gauche de la phrase/ de l'énoncé. En même temps, Rizzi distingue la topicalisation de l'articulation focus-présupposition, qui serait formellement équivalente en surface, mais dont la différence d'interprétation réside en la distribution de l'information : dans l'Exemple II-17, l'élément initial *your book* est interprété comme topic et comme information donnée (et le reste de la phrase est le 'comment'), tandis que dans Exemple II-18, l'élément initial *YOUR BOOK* correspondrait au focus, à l'information nouvelle, et le reste de la phrase est censé être 'présupposé'.

⁴⁰ L'exemple de base provient de Chomsky (1976 [1970]), dont Gundel discute les positions. La référence donnée dans l'article de Gundel est Chomsky (1971), qui correspond au même article reproduit ailleurs.

Exemple II-17 - Topicalisation (Rizzi, 1997 : 285)

Your book, you should give to Paul (not to Bill)

Exemple II-18 - Focus-présupposition (Rizzi, 1997 : 285)

YOUR BOOK you should give to Paul (not mine)

Les similitudes syntaxiques en surface seraient des traductions de structures profondes différentes, notamment une position pour 'topic' et une pour 'focus'. Dans le cadre de la grammaire transformationnelle, il est postulé une position syntaxique spécifique pour le topic ainsi que pour le focus en structure profonde : « topic-field » et « focus-field (1997 : 288). Il est précisé que ces champs sont présents seulement si nécessaire, ce qui laisse comprendre que Rizzi restreint topic et focus à des constructions spécialisées pour leur marquage explicite.

4.3.1.4 Gundel (1985)

Dans une version moins catégorique de ce genre d'approche syntaxique ou linéaire, Gundel (1985) part également d'une définition en termes de 'à propos' (cet aspect de son approche sera donc traité en section 4.3.3 ci-dessous), et mentionne aussi des critères de linéarisation, les énonçant toutefois sous forme de tendance plutôt que de régularité absolue. En effet, elle fait la différence entre les propriétés pragmatiques de thème/topic et ses propriétés structurelles, les premières étant susceptibles d'être universelles, les dernières dépendant de chaque système de langue. Ainsi, elle rappelle les travaux réunis dans le recueil de travaux intitulé « *Subject and Topic* » (Li, 1976), qui montrent que selon les langues, le principe de 'topic/comment' peut influencer plus ou moins sur la réalisation syntaxique de l'énoncé. Gundel note alors que, si dans une langue comme le japonais, le topic va toujours occuper la position initiale dans un énoncé, ce n'est pas nécessairement le cas dans une langue comme l'anglais. Toutefois, elle ajoute qu'il y a une tendance universelle forte pour les topics de se trouver en position initiale (ou initiale et finale), même dans des langues dans lesquelles la syntaxe est relativement indépendante de la dimension 'topic/comment', comme l'anglais (cf. 1985 : 93 sqq.).

4.3.1.5 Limites d'une définition sur critères formels

Il n'est guère difficile de montrer que la première position dans l'énoncé n'est pas un critère nécessaire, ni suffisant, pour déterminer le topic d'un énoncé : Reinhart (1981) par exemple argumente que dans l'exemple suivant, en fonction du contexte ou le syntagme nominal *Max* ou *Rosa* peut exprimer le topic :

Exemple II-19 – adapté de Reinhart (1981 : 56)

- (a) Max saw Rosa yesterday.
- (b) Who did Max see yesterday?
- (c) Did anyone see Rosa yesterday?

En effet, si Exemple II-19 (a) est prononcé en réponse à (b), (a) sera interprété comme étant à propos de Max, alors que si la question posée est (c) ; la réponse (a) ne pourra plus être interprété comme

étant à propos de Max. Nous redirons cependant, avec Gundel (1985), qu'il y a une tendance à trouver le topic en périphérie de l'énoncé, et notons que le travail avec les exemples inventés pour les besoins de l'analyse est souvent problématique : on doit se demander en effet si (a) sera effectivement produit comme réponse à (b) ou à (c). Il est probable qu'immédiatement après la mention du nom propre, la réponse fera appel plutôt qu'à un pronom, de sorte qu'une réponse naturelle à (c) par exemple sera plutôt quelque chose comme (d) ci-dessous :

(d) Max saw her yesterday.

Dans une visée plus théorique, Reinhart avance que des critères formels ne peuvent pas fournir une définition de la notion de topic, mais tout au plus une description de son marquage linguistique :

« The purely intonational approach, while consistent with the facts, does not define topics directly but rather the way they are marked linguistically. It leaves open the question under what discourse conditions a given expression would count as topic, and, consequently would be unstressed.

[...]

In a different framework, Givón (1976) argues that the topical hierarchy (which determines which NP is more likely to serve as a topic in a given sentence) is determined by four parameters, two of which are the human-nonhuman scale, and the definite-indefinite scale. Although these proposed parameters are no doubt helpful in the actual identification of the topics in given sentences, they do not explain, in and on themselves, what topics are (and they do not intend to do so). » (Reinhart, 1981 : 57)

4.3.2 Une définition du topic selon des critères informationnels

4.3.2.1 Le lien entre thème/topic et givenness

Jusqu'ici, nous avons vu que les définitions et descriptions de thème/topic recourent souvent à l'opposition entre information donnée et information nouvelle. Si dans les premiers travaux, chez Gabelentz (1869, 1891) ou Paul (1920 [1880]) par exemple, il s'agit de correspondances plutôt descriptives, dans d'autres travaux, le statut informationnel est élevé au rang de critère définitoire. Reinhart (1981) note ainsi que pour nombre de linguistes, le statut informationnel serait le seul facteur retenu pour définir ou identifier le topic d'un énoncé. Sans pouvoir donner un compte rendu exhaustif des travaux basés sur ces notions, nous allons tenter de retracer l'évolution et les affinements que cette dimension de la notion thème/topic a connus à travers quelques travaux exemplaires.

Notons en premier les travaux dans lesquels thème/topic est supposé être équivalent avec de l'information ancienne, connue ou donnée (*old, known* ou *given*)⁴¹, ou du moins suffisamment expliqué par la distinction *old/new*. Nous citons comme exemple une réflexion de Chafe (1970) :

⁴¹ voir ci-dessus la section 3 du présent chapitre.

« This distinction between old and new information is the principal phenomenon which underlies discussions of what have been called topic and comment, or theme and rheme [...] » (1970 : 211)

Chez Givón (1992), nous trouvons une définition qui va dans le même sens, et qui met en équivalence information ancienne (*old*) et topic :

« The chunks of old, redundant ('topical') information in the clause serve to *ground* the new information to the already-stored old information. Cognitively, they furnish the *address* or *label* for the *storage locus* ('file') in the episodic memory. »⁴²
(1992 : 9)

Toutefois, ces positions fortes se trouvent quelque peu plus modulées ailleurs. Ainsi, Chafe nuance dans un article ultérieur : « there is no necessary correlation of subject status with givenness, or for that matter of nonsubject status with newness »⁴³ (1976 : 48).

4.3.2.2 Critique de l'association entre thème/topic et givenness

D'autres travaux séparent les dimensions thème/topic et structuration informationnelle. Comme mentionné précédemment, Halliday (1967) distingue entre 'thème' et 'information donnée', le dernier étant le principe qui lie un énoncé au discours précédent, et le premier exprimant l'orientation actuelle du locuteur :

« [...] there is in the unmarked case (i.e. if the information structure is unmarked) an association of the theme with the given, the two are independent options (cf. Firbas, 1964). The difference can perhaps be best summarized by the observation that, while 'given' means 'what you were talking about' (or 'what I was talking about before'), 'theme' means 'what I am talking about' (or 'what I am talking about now'); and, as any student of rhetoric knows, the two do not necessarily coincide.
The information systems, in other words, specify a structural unit and structure it in such a way as to relate it to the preceding discourse; whereas thematization takes a unit of sentence structure, the clause, and structures it in a way that is independent of what has gone before. » (Halliday, 1967 : 211-212)

Par ailleurs, Halliday rejette explicitement le terme de *topic* en raison du cumul observé dans la littérature entre ce qui est thématique et l'information connue :

⁴² Notons par ailleurs que le cumul fonctionnel opéré ici par Givon entre information déjà connue et *label* pour la nouvelle information est précisément ce qui motive Halliday pour différencier information donnée et thème ; nous y reviendrons plus loin dans cette section.

⁴³ Nous rappelons au lecteur que pour Chafe, dans cet article, *subject* semble correspondre à ce que d'autres appellent *topic*, cf. section 4.3.1.2).

« [...] 'topic' and 'comment' are however avoided because they have tended to be used in a way which conflates what are here regarded as distinct functions, with 'topic' meaning both 'given' and 'theme'. » (1967 : 200)

Les deux dimensions thématique et informationnelle s'articulent alors dans un système complexe dans l'unité intonative, qui correspondra souvent, à l'oral, à un énoncé :

« While therefore the given - new structure is not itself realized by the sequence of elements, and the focus of information may fall anywhere in the information unit, the partial congruence between this variable and the one which is in fact realized by the sequence of elements, that of theme - rheme, together with the partial congruence between clause and information unit, results in a tendency towards a left to right form of organization in the information unit with given, if present, preceding new. » (1967 : 205)

De même, Reinhart (1981) démontre, à l'aide de contre-exemples, que l'association entre *topic* et information donnée n'est pas systématique, et qu'en conséquence, la notion de *topic* ne saurait être définie en termes de *givenness*. En effet, elle montre que ce critère ne constitue ni une condition suffisante, ni nécessaire. D'un côté, puisque tout ce qui est information donnée dans un énoncé ne constitue pas nécessairement le *topic* de cet énoncé, ce critère n'est pas suffisant pour définir la notion *topic*. D'un autre côté, ce n'est pas une condition nécessaire non plus. Un des cas qu'elle discute concerne des expressions indéfinies spécifiques :

Exemple II-20 : Specific indefinite topic – tiré de Reinhart (1981 : 66)

When she was five years old, *a child of my acquaintance* announced a theory that she was inhabited by rabbits

L'auteure note que l'expression en italique n'est pas de l'information donnée, mais doit néanmoins être interprétée comme le *topic* de cet énoncé, selon sa définition en termes de 'aboutness' (cf. section 4.3.3 ci-dessous pour une présentation détaillée) :⁴⁴

« By the old-information criterion for topichood the underlined indefinite expressions [...] cannot be topic expressions, since they obviously represent new information, i.e. the existence of these referents cannot be assumed to be in the hearer's immediate awareness. (Otherwise a definite expression would have been used.) But by the intuitive aboutness criterion, they can be understood as asserting something about the child [...]. » (Reinhart, 1981)

L'auteure ajoute toutefois que des analyses empiriques comme celle de Ochs Keenan & Schieffelin (1976a) ont montré une tendance pour le *topic* de représenter de l'information ancienne.

⁴⁴ Notons également que l'expression est ancrée et rendue identifiable par l'emploi du pronom possessif *my*, cf. la notion d'ancrage en section 3.2.1 ci-dessus.

Elle souligne au même temps que la question de la topicalité ne peut pas être définie directement en termes de statut informationnel d'un référent. Dans l'exemple ci-dessous, si les énoncés sont interprétés comme étant construits à propos du référent Felix, en effet, une contradiction surgit :

Exemple II-21 - (Reinhart, 1981 : 72)

- (a) A : Who did Felix praise?
B : Felix praised MAX.
- (b) A : Who did Felix praise?
B : Felix praised HIMSELF.

Si le statut de topic devait être défini en fonction du statut informationnel du référent, il n'est pas clair pourquoi en Exemple II-21 (b) l'expression nominale *Felix* est topic, et non pas le pronom *himself*, les deux représentant le même référent et de l'information donnée. Gundel (1985 : 90) considère néanmoins que ce dernier argument est moins convaincant, étant donné qu'il a déjà été montré ailleurs que différentes acceptions de *old/new* existent, et que ce qui est en cause ici n'est pas tant le statut informationnel du référent lui-même, mais sa relation avec le reste de la proposition (cf. les remarques sur Kuno (1972) en section 3.1.1 ci-dessus, qui proposait déjà une analyse différenciée).

Lambrecht (1994 : 131) insiste également sur la distinction à faire entre « topic referent » et « topic expression », permettant de contourner l'apparente contradiction de l'exemple cité par Reinhart (1981). Ainsi, ce qui caractérise le topic n'est pas le statut attentionnel d'un référent, mais la relation établie entre ce référent et une prédication (Lambrecht emploie le terme proposition). Blasco-Dulbecco aussi considère qu'« un élément donné ou nouveau dans un énoncé concerne ce que l'on appelle la structure informationnelle de la phrase. En revanche, d'après les critères usuels, le thème, en tant qu'élément sélectionné pour en dire quelque chose, ne relève plus de la structure informationnelle de la phrase. Il n'est donc pas justifié de mettre sur le même plan ces deux notions car il s'agit de deux caractéristiques complémentaires et indépendantes, et on risque une certaine confusion lorsque l'on tente de les faire coïncider. » (Blasco-Dulbecco, 1999 : 64-65).

Cette position, que Lambrecht ou Blasco-Dulbecco ne sont pas les premiers à défendre, caractérise un déplacement des investigations sur thème/topic dans la littérature d'une approche purement informationnelle vers des approches plus fonctionnalistes.

4.3.3 Une définition pragmatique du topic

Définir la notion de thème/topic comme *ce dont on parle* est une approche bien plus vague que les tentatives de définition décrites ci-dessus. En même temps, elle présente l'avantage de limiter le risque d'amalgame entre fonctions langagières qui peuvent se manifester par les mêmes propriétés formelles, sans être de nature identique (cf. les développements sur la linéarité de la parole et la dimension informationnelle ci-dessus). Certains auteurs distinguent explicitement ces niveaux : nous avons discuté la position de Halliday (1967) en section 4.3.2 ci-dessus, qui insiste sur la nécessité de

séparer les niveaux d'organisation informationnelle et thématique, de distinguer information donnée et thème. Kuno (1972, 1978, 1981) défend un point de vue similaire : Dans la tradition de la « Functional Sentence Perspective » de l'école de Prague, il analyse les liens entre les dimensions thème/rhème et old/new information pour le japonais et l'anglais. Comme nous l'avons déjà exposé en section 3.1.1 ci-dessus, il fait une distinction entre deux acceptions de 'old', la première correspondant au fait qu'une expression soit anaphorique, déjà mentionnée dans le discours précédent ou entrée dans le registre permanent, et la deuxième étant exprimée en termes de « predictability ». Kuno distingue ces dimensions d'anaphorique/non-anaphorique et *old (predictable)/new (unpredictable)* de la dimension de thème. Il poursuit en rejetant explicitement l'amalgame entre ces différentes dimensions (Kuno, 1978 : 283-284).

Quant à la définition de la notion de thème, Kuno s'en acquitte en passant, sans élaboration particulière, comme suit : « The concept of "theme" cannot be given any precise formulation. I can only say that the theme is what the rest of the sentence is about. » (1978 : 283). Ce principe définitoire en termes d'*aboutness* est parmi les plus usités dans des approches pragmatiques de la question du thème/topic. Gundel (1985) note en effet que la fonction principale du thème/topic, celle d'établir la pertinence d'un énoncé est consensuelle pour la plupart des auteurs. Ce seraient plutôt les propriétés structurelles et pragmatiques de cette catégorie, que nous avons exposées dans les sections précédentes, qui font l'objet du débat.

Si l'approche la plus souvent citée dans le cadre d'une telle définition est celle de Reinhart (1981)⁴⁵, pour l'auteure elle-même l'origine de cette définition est à chercher dans les travaux de Strawson (1964),⁴⁶ que nous allons présenter dans ce qui suit.

4.3.3.1 Strawson (1964), la référence et les valeurs de vérité

L'article de Strawson est davantage de nature logico-philosophique que linguistique, mais touche néanmoins à des points très pertinents pour le linguiste. L'objet de ses réflexions sont la référence et la controverse autour de la question des valeurs de vérité (Frege, 1892 notamment). Un énoncé peut, en termes de logique formelle, être soit vrai, soit faux, en fonction de son adéquation avec le monde réel : l'énoncé *Jean est professeur de linguistique* sera vrai si Jean possède effectivement cette qualification, et faux si, au contraire, il est, disons, maçon. La controverse surgit alors lorsqu'il s'agit d'apprécier des énoncés qui contiennent une expression référentielle qui ne réfère pas : dans l'exemple célèbre *the present king of France is bald* l'expression *the present king of France* ne réfère pas, selon Strawson, puisqu'une personne telle qu'un actuel roi de France n'existe pas. Deux positions s'opposent alors : soit l'on peut dire, avec Russell par exemple, de cet énoncé qu'il est aussi faux que l'exemple précédent sur Jean, soit, en suivant Quine (Quine, 1960), on dira que tel énoncé

⁴⁵ Parfois cité comme Reinhart (1982), par exemple dans Gundel (1985) : il s'agit toutefois du même article.

⁴⁶ cf. aussi la synthèse dans Gundel (1985).

n'est ni vrai ni faux : c'est ce que Quine appelle le *truth-value-gap*. À partir de cette controverse donc, Strawson se livre à l'examen de quelques exemples qui lui permettront de dire qu'un énoncé ne peut être interprété lorsque l'expression qui manque de référer correctement en constitue le *topic*, et qu'en ce cas, il y aurait effectivement un *truth-value-gap*. En revanche, lorsque l'expression qui manque de référer n'est pas le *topic* de l'énoncé, mais s'intègre dans le commentaire ou propos (« absorbed into a predicate-term which is attached to the innocent referring expression to make up the statement as a whole » (Strawson, 1964 : 111)), l'énoncé peut être évalué comme vrai ou faux. Ainsi, selon l'auteur, l'exemple (a) ci-dessous est construit avec l'expression *the Exhibition* comme *topic*, alors que le *topic* de (b) est *the king of France*. L'énoncé (a) peut alors être interprété, pour peu qu'une *Exhibition* se soit effectivement tenue, et l'assertion serait alors simplement fautive en vertu de la non-existence d'un roi de France, alors que cette même non-existence empêcherait l'énoncé (b) d'être interprété, qui ne serait alors ni vrai ni faux.

Exemple II-22 – tire de Strawson (1964)

- (a) The Exhibition was visited yesterday by the king of France.
- (b) The king of France visited the Exhibition yesterday.

Ce qui nous intéresse particulièrement ici n'est pas tant la controverse autour des valeurs de vérité, mais les considérations suivantes :

a. Les expressions référentielles servent la fonction d'identification du référent

Strawson s'intéresse donc aux cas dans lesquels une expression référentielle est employée dans un énoncé de façon adaptée, afin de permettre à l'interlocuteur d'identifier le référent. Cette réflexion rappelle les questions traitées en section 3 ci-dessus, et mène effectivement Strawson à la considération suivante :

b. Principle of Knowledge

L'auteur lie la question de l'identification de la référence à un principe de connaissance présumée : l'observation que la communication est constituée non seulement d'informations nouvelles, mais contient également des bouts de savoir déjà partagés avec l'interlocuteur, auxquels le locuteur fera référence moyennant une expression référentielle adéquate :

« When people talk to each other they commonly and rightly assume a large community of identifying knowledge of particular items. Very often a speaker knows or assumes that a thing of which he has such knowledge is also a thing of which his audience has such knowledge. Knowing or assuming this, he may wish to state some particular fact regarding such a thing, e.g. that it is thus-and-so; and he will then normally include in this utterance an expression which he regards as adequate, in the circumstances of utterance, to indicate to the audience *which* thing it is, of all the things in the scope of the audience's identifying knowledge, that he is declaring to be thus-and-so. » (Strawson, 1964 : 99)

c. Principle of Relevance

Les observations de Strawson aboutissent à la considération qu'en général, la communication n'est pas l'échange fortuit d'informations non connectées entre elles, mais que les locuteurs « on the contrary intend in general to give or add information about what is a matter of standing or current interest or concern » (1964 : 115). C'est cette dernière notion de *current interest* qui sera importante pour la définition pragmatique du topic. Strawson associe ce terme avec ce qu'il appelle *topic* et aussi *subject* (et il entend par là justement non pas le sujet logique ou grammatical, mais bien *ce dont on parle* : cette notion est donc très proche du sujet psychologique de Gabelentz (1869, 1891) :

« Statements, or the pieces of discourse to which they belong, have subjects, not only in the relatively precise senses of logic and grammar, but in a vaguer sense with which I shall associate the words 'topic' and 'about'. [...] the idea of the topic or centre of interest of a statement, the idea of what a statement could be said, in this sense, to be about. » (Strawson, 1964 : 114)

Strawson termine son analyse en notant que si les principes de *knowledge* et celui de *relevance* se chevauchent souvent, ce n'est pas nécessairement le cas.

Comme le note Reinhart, la notion de *topic* n'est donc pas réellement définie non plus par Strawson, mais seulement circonscrite par des tournures comme *center of interest* ou *about*. En revanche, l'auteur donne des critères permettant d'identifier le topic dans un énoncé donné, qui sera connu comme le test de la paraphrase ou de la reformulation, et qui revient à enchâsser l'énoncé en question dans une description de l'acte du langage accompli : pour reprendre les énoncés de l'exemple Exemple II-22, les paraphrases correspondantes sont données en (a1) et (b1) :

Exemple II-23 – Paraphrase pour identifier le topic selon Strawson (1964 : 117)⁴⁷

- (a) **The Exhibition** was visited yesterday by *the king of France*.
- (a1) He/ She was describing by *whom* **the exhibition** was visited yesterday.
- (b) **The king of France** visited *the Exhibition* yesterday.
- (b1) He/ She described *what kind of event* **the king of France** attended to yesterday.

Selon Strawson, le topic (en gras dans l'exemple) apparaîtra dans la paraphrase, alors que ce qui n'est pas topic sera remplacé par le pronom interrogatif (en italiques dans l'exemple ; cf. aussi les positions de Chomsky en section 4.3.1.3 ci-dessus). Nous reviendrons en CHAPITRE IX-1.2.2 sur les différents critères et tests permettant d'identifier le topic.

4.3.3.2 Restrictions et critères du thème/topic chez Reinhart (1981)

Reinhart (1981) propose, quant à elle, de fournir une définition plus robuste de la notion de *pragmatic aboutness*. Ayant montré que topic et information donnée ne coïncident pas

⁴⁷ Dans les paraphrases notées ici, nous avons marqué comme topic seulement l'expression référentielle. Strawson reste quelque peu ambigu sur ce point : son intérêt principal sont les expressions référentielles, porteuses d'une présupposition d'existence, mais de fait, ce qui persiste dans la paraphrase est d'une étendue plus large, et pourrait ainsi correspondre à la présupposition de Chomsky : *The exhibition was visited by somebody yesterday* et *The king of France visited something yesterday*.

nécessairement, elle propose d'envisager le topic alors non pas comme effet du discours précédent, mais en fonction des effets qu'a le topic lui-même sur la suite du discours.⁴⁸ L'auteure évoque pour cela la métaphore d'un catalogue de bibliothèque, et compare le topic avec les entrées de catalogue organisés non pas par titre, mais par sujet : chaque livre sur ce sujet correspond alors à une proposition à propos de ce topic. Pour Reinhart, « sentence-topics, within this view, are one of the means available in the language to organize, or classify the information exchanged in linguistic communication - they are signals for how to construct the context set, or under which entries to classify the new proposition. »

Dans son article, Reinhart énonce des restrictions auxquelles une expression référentielle doit se conformer pour pouvoir être interprété comme topic : ainsi, avec Strawson (1964), elle note que « only sentence-topics carry existential presuppositions » (Reinhart, 1981 : 70). Il s'ensuit entre autres que des descriptions nominales avec quantifieurs existentiels (*un, certains, quelques*, etc. : lorsqu'ils doivent être interprétés sous forme de « il existe au moins un X tel que Y », donc en dehors d'une interprétation spécifique ou générique) ne peuvent être interprétées comme topic. Reinhart donne l'exemple suivant : *Rosa's friends should call a doctor*, où les expressions *Rosa* ou *Rosa's friends* pourraient, en fonction du contexte, être interprétés comme topic, mais non pas l'expression *a doctor*. À l'inverse, elle note que lorsqu'un énoncé contient une expression référentielle disloquée ou topicalisée, celle-ci sera obligatoirement interprétée comme le topic (à noter que la dislocation d'un SN indéfini est rare, et souvent à référence générique dans ce cas : *un cheval, ça mange de l'herbe*).

L'auteure avance des critères supplémentaires qui devraient permettre d'identifier le topic d'un énoncé donné, mais ces critères sont formulés en termes de préférences que nous avons déjà eu l'occasion de discuter dans ce chapitre : ainsi, référent donné dans le discours aurait plus de chances d'être encodé comme topic qu'un référent nouveau, de même qu'il y aurait une préférence pour interpréter le sujet grammatical comme topic de l'énoncé.

4.3.3.3 *Firbas et le Dynamisme Communicatif*

Nous avons mentionné en entrée de ce chapitre l'héritage des philosophes de l'antiquité qu'est la bipartition de l'énoncé. Jusqu'ici, les définitions de thème/topic que nous avons présentées soit s'y sont explicitement conformés, soit ne discutent pas cette analyse.

Certains auteurs analysent l'énoncé en topic/comment, d'autres en focus/présupposition, en fonction de l'élément (thème/topic ou focus) choisi comme point de départ pour l'analyse. Ces approches ne s'intéressent donc pas (ou alors pas de manière centrale) à l'organisation interne des deux grandes

⁴⁸ Cette position rappelle d'ailleurs la distinction que fait Halliday entre la dimension informationnelle et celle de thème/topic, liant la première au discours précédent et la seconde à l'orientation actuelle du discours, cf. la citation de Halliday en section 4.3.2.2 ci-dessus.

parties ainsi identifiées. D'autres approches vont davantage s'intéresser aux réalisations des phénomènes extrêmes (*theme et rheme proper/focus*), délaissant le 'reste' de l'énoncé.

Firbas (i.a. 1964, 1992) développera la notion de Dynamisme Communicatif qui permettra, dans le cadre de la Functional Sentence Perspective, une analyse de la phrase selon une échelle plus fine.⁴⁹ Le dynamisme communicatif est compris comme une orientation de chaque énoncé ou proposition vers l'élément qui accomplit (« consummates or completes ») son but communicatif (Firbas, 1992 : 6). Cette dimension peut correspondre à l'arrangement linéaire des éléments dans l'énoncé, mais ce n'est pas nécessairement le cas. Pour Firbas, le thème correspond à l'unité qui représente le plus faible degré de dynamisme communicatif, tandis que le rhème, à l'autre extrême de l'échelle, correspond au degré le plus fort du dynamisme communicatif. Entre les deux peuvent exister des éléments transitionnels qui ne sont ni thème, ni rhème. L'auteur ajoute que « a thematic element – if present - will constitute the starting point of this arrangement. It will be the element carrying the lowest degree of CD [i.e., Communicative Dynamism; *notre remarque*] within the theme. In the absence of the theme, the starting point is constituted by the element carrying the lowest degree of CD within the non-theme. » (1992 : 73). Firbas conclura ses développements sur les éléments thématiques et non-thématiques en précisant comment il se positionne par rapport au critère de *aboutness* :

« It may have been observed that in delimiting the theme I have not had recourse to the notion of 'aboutness'. This does not mean that I do not subscribe to the view that the theme expresses what the message conveyed is to be about. But the 'aboutness' feature is not the starting point of my delimitation of the theme; it is its outcome that singles out the elements that convey the theme and bear the 'aboutness' feature. » (1992 : 73)

4.3.3.4 La définition de thème/topic chez Gundel et Lambrecht

En section 4.3.1.4, nous avons déjà évoqué les positions de Gundel (1985) par rapport aux caractéristiques formelles du thème/topic, et nous avons annoncé que malgré tout, sa définition de topic était à classer plutôt parmi les approches présentées dans la présente section. Gundel (1999) situera son approche dans les termes suivants :

⁴⁹ cf. aussi Lyons (1977b : 509) :

« So far we have taken the view that theme and rheme are complementary, in the sense that whatever is not thematic is rhematic and whatever is not rhematic is thematic. Firbas (1964, 1972) takes a different view. He operates with a notion of communicative dynamism (CD), defined in terms of the degree to which an expression advances, or fails to advance, the process of communication. »

« I agree with Reinhart that sentence topic is a specifically linguistic phenomenon. But I will propose in this paper that it is not primarily a pragmatic or discourse phenomenon as Reinhart and others have assumed. It is an integral part of the semantic/conceptual representation of natural language sentences, which is encoded (though not always unambiguously) by their morpho-syntactic and/or phonological form. [...] The important question then isn't whether some particular linguistic phenomenon has pragmatic effects or not, but which of its properties are determined by the grammar and which can be derived from more general cognitive and communicative principles. » (J. K. Gundel, 1999 : 185)

Cette approche, qui situe la notion de topic à l'interface entre syntaxe et pragmatique, est formalisée par une distinction entre le topic pragmatique - en tant que relation entre locuteur et énoncé par rapport à un contexte donné -, et le topic syntaxique - en tant que relation entre constituant et énoncé - comme suit :

« I will make a distinction here between pragmatic topic, a relation that holds between a speaker and a sentence relative to a context, as defined in (5) and syntactic topic, a relation that holds between a constituent and a sentence which contains that constituent, and which is defined directly on syntactic structures, as in (6).

(5) *Pragmatic Topic*. Def.:

An entity, E, is the pragmatic topic of a sentence, S, iff S is intended to increase the addressee's knowledge about, request information about or otherwise get the addressee to act with respect to E.

(6) *Syntactic Topic*. Def.:

A constituent, C, is the syntactic topic of some sentence, S, iff C is immediately dominated by S and C is adjoined to the left or right of some sentence S' which is also immediately dominated by S. »

(J. K. Gundel, 1985 : 86; les italiques sont de l'auteure)

L'auteure précise qu'un constituant qui, structurellement, doit être interprété comme le topic d'un énoncé, est toujours également le topic pragmatique de cet énoncé, mais l'inverse n'est pas vrai. Le topic syntaxique est en fait rien d'autre que des constructions spécialisées dans l'encodage du topic : Pour Gundel (1988), il s'agit de la dislocation (à gauche ou à droite) ainsi que du hanging topic (« double subject » dans son article ; (1988 : 223-224), exemplifiés ci-dessous :⁵⁰

Exemple II-24 – Syntactic-topic constructions ; adapté de Gundel (J. K. Gundel, 1988 : 224)

- (a) **Die Frau da, sie** kommt aus Berlin (Dislocation à gauche)
the woman there she comes from Berlin
Cette femme là elle est de Berlin.
- (b) **He's** shrewed, **that one** (Dislocation à droite)
- (c) **My work,** I'm going crazy (Hanging Topic)

Cette définition de thème/topic présente l'avantage de définir le topic (pragmatique) par rapport au référent, et non pas par rapport à un constituant de l'énoncé. Il est ainsi possible de l'appliquer

⁵⁰ Le lecteur trouvera une description détaillée de ces constructions dans le CHAPITRE IV.

également à des énoncés dont le topic n'est pas exprimé par une expression marquée, comme les pronoms, ou encore des cas où le topic n'est pas ouvertement exprimé : dans l'exemple (a) suivant, tiré de Gundel (1988), en fonction du contexte deux lectures de l'énoncé sont possibles. Gundel explique que dans la première interprétation, *Marcos* exprime le topic, *resigned* le comment (ou focus, selon les terminologies). Dans la seconde lecture, l'énoncé entier représenterait le comment, le topic correspondant alors à un cadre plus large, comme 'la situation politique dans les Philippines'. Gundel justifie ensuite la distinction faite entre topic pragmatique et topic syntaxique, puisque ces derniers sont toujours nécessairement interprétés comme topic pragmatique, et la deuxième lecture, où tout serait comment, n'est pas possible avec l'exemple (b) :

Exemple II-25 – Topic syntaxique et topic pragmatique, adapté de Gundel (1988 : 211)⁵¹

- (a) *Marcos resigned*
- (b) *Marcos, he resigned*

Notons enfin la définition de Lambrecht, qui fait également une distinction entre la catégorie pragmatique *topic* et son expression linguistique, mais dont la définition est moins syntaxique :

« TOPIC: A referent is interpreted as the topic of a proposition if in a given situation the proposition is construed as being about this referent, i.e. as expressing information which is relevant to and which increases the addressee's knowledge of this referent.

TOPIC EXPRESSION: A constituent is a topic expression if the proposition expressed by the clause with which it is associated is pragmatically construed as being about the referent of this constituent. » (Lambrecht, 1994 : 131)

La définition pragmatique de topic semble quelque peu plus restrictive que celle de Gundel, puisqu'elle pourra s'appliquer plus difficilement à des énoncés autres que des assertions, dans la mesure qu'il sera difficile de dire qu'une question ou un ordre ajoutent à la connaissance de l'interlocuteur par rapport au référent en question. D'un autre côté, la notion de topic *expression* explicite la distinction opérée par Gundel, tout en permettant de traiter toute expression (explicite ou non) du topic, et non seulement les constructions spécialisées. Nous montrerons par ailleurs dans le CHAPITRE IV-1.4.1.5, que la généralisation selon laquelle ces constructions expriment toujours le topic pragmatique, bien qu'elle soit très forte, connaît des exceptions.

4.3.3.5 *Topic d'à propos et topic cadratif*

Une dernière distinction est cependant nécessaire, celle entre le topic d'à propos au sens strict, et ce qui a été appelé *stage topic*, topic cadratif ou scénique. Nous pouvons nous baser pour cela sur la synthèse proposée par Horváth (2018 : 43-44) à ce sujet, ainsi que sur l'article de Prévost (2003). Rappelons la caractérisation donnée par Chafe, déjà citée en section 4.3.1.2, pour qui le topic cadratif

⁵¹ Dans les exemples de Gundel (1988), l'italique représente l'accentuation principale de l'énoncé (*primary stress*)

« sets a spatial, temporal, or individual framework within which the main predication holds » (Chafe, 1976 : 50). D'autres auteurs, par exemple Gundel (1989 [1974]) ou Erteschik-Shir (1997, 2007) limitent le topic cadratif aux seuls paramètres spatio-temporels. L'origine de cette notion était alors de pouvoir attribuer un topic à ces énoncés qui ne semblent pas en avoir, les énoncés théétiques, qui sont entièrement en focus :

Exemple II-26 – notre traduction des exemples donnés par Erteschik-Shir (2007 : 17)

- (a) Il pleut.
- (b) Il y a un chat devant la porte.

Ces énoncés ne sont pas *à propos* de quelque chose, mais afin de les évaluer correctement, l'allocutaire doit avoir recours au cadre spatio-temporel du discours, l'ici et maintenant. Ce cadre peut donc être implicite, et s'il est exprimé, correspond typiquement à des circonstants :

Exemple II-27 – (Horváth, 2018 : 43)

- (a) **Ce matin** j'ai rencontré Marie. Elle n'a pas l'air très en forme. (Prévost 2003, 56)
- (b) **Sur le Bassin d'Arcachon**, on continue à trouver des galettes de fuel. Il en revient à chaque grande marée. (Prévost 2003, 56)
- (c) **Au 17^e siècle**, la condition paysanne était rude. (Le Goffic 1993, 463)
- (d) **Physiquement**, Pierre va bien. (D'après Jacobs 2001, 655)

Ces éléments constituent donc le cadre dans lequel l'énoncé peut être évalué, mais ils ne constituent pas l'*à propos* de cet énoncé. Il en va autrement pour le type de cadre supplémentaire assumé par Chafe pour les *chinese style topics*, qu'il appelle « individual framework » (1976 : 50). Pour cet auteur, il ne s'agit pas non plus d'un topic d'*à propos*, mais Prévost (2003 : 3-4) réfute cette analyse et propose, pour la notion de topic d'*à propos* « qu'elle puisse se réduire à une relation pragmatique de pertinence », formulée comme « ce à propos de quoi il est pertinent de dire quelque chose ». Elle soulève la proximité de cette notion avec le topic non lié de Lambrecht, traité aussi comme dislocation sans reprise ou hanging topic dans d'autres approches (voir notre CHAPITRE IV pour une discussion détaillée) :

Exemple II-28 - (Prévost, 2003 : 4)

Le métro, avec la carte orange tu vas partout.

Prévost distingue ainsi le « cadre énonciatif (notionnel/thématique) *par rapport auquel* il est pertinent d'énoncer la relation prédicative qui suit » du « cadre (spatial ou temporel) pour l'énoncé, *dans lequel* la relation prédicative qui suit est pertinente » (Prévost, 2003 : 6). Malgré cette première distinction, l'auteure estime que les notions de topic et de cadre ne s'excluent pas totalement en ce qui concerne la dimension notionnelle/thématique. Elle montre en outre que même pour les cadres spatiaux et temporels, contrairement à la distinction posée jusqu'ici, une relation d'*à propos* n'est pas exclue. Cela dépend selon elle de la suite du discours. Pour les deux énoncés suivants,

Exemple II-29 - (Prévost, 2003 : 7)

La nuit, tous les chats sont gris,
Au 17^{ème} siècle, la condition paysanne était rude,

l'auteure suggère que « [si] le locuteur poursuit son discours sur les chats, ou sur la condition paysanne, ceux-ci constituent le topique principal de l'énoncé » (2003 : 8). Une continuation du discours sur la nuit ou sur le 17^{ème} siècle, au contraire, imposerait une interprétation de ces cadres spatio-temporels également comme topic de leur énoncé. Ainsi, « seul le contexte permet de trancher en faveur de l'interprétation cadrative ou topicale » (2003 : 15). Notons que dans nos analyses, nous nous limiterons aux topics de l'énoncé, et laisserons de côté les topics cadratifs.

4.3.4 Topic comme notion scalaire : une échelle universelle de la topicalité (Givón, 1983)

Il est curieux de constater que dans les travaux de Givón, il n'est pas aisé de trouver une définition claire de ce qu'il entend par *topic*. La formulation *topic/participant* laisse déjà entendre que sa définition n'identifie pas a priori une fonction unique dans la phrase : en effet, Givón (1983 : 6) estime que dans un exemple tel que le suivant, il y aurait simultanément trois topics différents, « perhaps hierarchized by degree », dont le topic principal serait le sujet en tant que topic grammaticalisé :

Exemple II-30 – Topics multiples dans l'approche de Givón (1983 : 6)

- (a) *John gave the book to Mary*
- (b) *John gave Mary the book*

Pour Givón, topic n'est donc pas une fonction catégorielle établie, discrète, et il propose plutôt une notion graduelle de topicalité. Un référent peut donc être plus ou moins topical. Si pour Givón « the clause ('sentence') is the basic information processing unit in human discourse, l'étude de la topicalité devra se faire dans le cadre plus large de paragraphes thématiques (1983 : 7). Lorsque Givón explique de quelle manière sont imbriquées la continuité thématique, celle des topics ainsi que celle des actions, il devient transparent que Givón utilise le terme thème au niveau du discours, et celui de topic au niveau de l'énoncé :

« Thematic continuity is the overall matrix for all other continuities in the discourse. It is the hardest to specify, yet it is clearly and demonstrably there. Statistically, it coincides with topic and action continuity to quite an extent within the thematic paragraph. The thematic paragraph is by definition about the same theme. Most commonly it also preserves topic and action continuity. »
(1983 : 8)

Il nous semble finalement que topic pour Givón est défini pragmatiquement, en termes de *aboutness*, même si cela n'est pas dit explicitement (mais le mot *about* est employé dans la citation ci-dessus pour définir et caractériser le thème). En effet, lorsque Givón expose un point central des questions de recherche posées dans les différentes études réunies dans *Topic Continuity in Discourse* (Givón,

1983), il emploie la métaphore des fiches d'un registre, sous forme desquelles un topic doit être enregistré et qui permettra ensuite de traiter correctement des nouvelles informations :

« the *degree of difficulty* that speakers/hearers may experience in *identifying a topic* in discourse, i.e. in filing it appropriately in their internal register, so that predications or new information transmitted about those topics would in turn be *addressed* correctly » (1983 : 11)

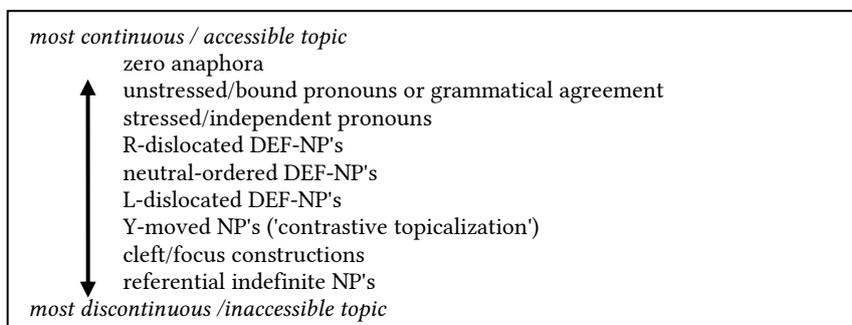
Cette métaphore est la même, avec des variations terminologiques, que celle du catalogue de bibliothèque donnée par Reinhart (1981) et présentée en section 4.3.3.2 ci-dessus, ou encore l'image du *hitching post* de Chafe (cf. section 4.3.1.2 ci-dessus), qui séparent donc l'information en ce dont on parle et ce que l'on en dit.

La particularité de l'approche de Givón est alors qu'il n'aura pas recours à cette métaphore comme définition intuitive pour identifier le topic. Il explique que si l'encodage linguistique de topicalité est basé sur les assomptions d'un locuteur sur la facilité avec laquelle un interlocuteur pourra identifier le topic, ces considérations psychologiques ne sont pas directement accessibles au linguiste qui analyse un texte donné : toute expression référentielle est traitée comme topic potentiel, et une batterie assez impressionnante de mesures en détermine le degré de topicalité.⁵²

L'auteur combine ainsi les caractéristiques morphosyntaxiques avec des considérations discursives pour établir la topicalité d'une expression référentielle. Les mesures discursives concernent la distance entre une expression référentielle et la dernière mention de son référent (*referential distance* ou *look-back*), l'ambiguïté potentielle provoquée par des référents concurrents dans le discours précédent (*potential interference* ou *ambiguity*), ainsi qu'une troisième mesure qui, elle, n'est pas orienté vers le discours précédent, mais subséquent, et qui mesure pendant combien de propositions l'élément en question continue d'être mentionné sans interruption.

Pour ce qui est des caractéristiques morfo-syntaxiques, Givón propose plusieurs échelles, dont la plus connue est celle des formes des expressions référentielles :

Figure II-7 - Distribution des expressions référentielles sur l'échelle de topic accessibility (Givón, 1983 : 17)



⁵² Givon (1983 : 12) indique par ailleurs qu'il emploie le terme *continuity* avec une grande proximité conceptuelle des notions d'accessibilité ou identifiabilité, puisque ce qui est continu serait plus prévisible et ainsi plus facile à interpréter.

Cependant, Givón poursuit son argumentation en admettant que cette échelle ne saura prétendre à l'universalité, notamment parce qu'elle est trop complexe : l'établissement de plusieurs échelles rendant compte chacune d'un domaine de d'encodage linguistique serait plus adéquat, selon l'auteur. Il propose différentes échelles qui traduisent la continuité/accessibilité topicale selon les domaines du poids phonologique, de l'intonation, de l'ordre des mots, du rôle sémantique ou encore de la fonction syntaxique. Pour le détail de ces échelles, nous renvoyons le lecteur au chapitre de Givón, mais nous pouvons dire que ces échelles résument très bien les observations diverses d'autres auteurs que nous avons pu rapporter dans les sections précédentes : ainsi, selon Givón, une expression traduisant un haut degré de topicalité sera d'un poids phonologique très réduit (forme Zero dans le cas le plus extrême), non accentuée, sujet grammatical et exprimant l'agent de l'action. En ce qui concerne l'ordre des mots, il sera nécessaire de développer un petit peu. L'échelle donnée par Givón est la suivante :

Figure II-8 – Ordre des mots dans les langues à ordre rigide (Givón, 1983 : 19)

R-dislocation > neutral word-order > L-dislocation
--

L'auteur précise que cette échelle est valable pour des langues dont l'ordre des mots est rigide comme le japonais ou l'anglais. Elle est certainement pertinente aussi pour le français. Des éléments disloqués à droite traduiraient le degré d'accessibilité le plus fort, ceux disloqués à gauche le degré le plus faible. L'ordre de mots neutre se trouvant entre les deux, cette échelle relativise quelque peu le postulat selon lequel topic correspondrait à la position initiale, puisque position initiale neutre et dislocation à gauche peuvent être différenciées.

Pour les langues à l'ordre de mots plus flexible, Givón propose l'échelle suivante :

Figure II-9 – Ordre des mots dans les langues à ordre flexible (Givón, 1983 : 19)

(a) VS > SV
(b) VO > OV

Enfin, Givón note que les deux échelles données en Figure II-8 et Figure II-9 ci-dessus seraient essentiellement les mêmes, et traduiraient in fine la position relative du topic et du comment. Si l'on intègre l'échelle de la taille phonologique, l'on aboutirait, selon l'auteur, à l'échelle suivante, qui prend en compte le fait que dans le cas de très forte accessibilité du topic, un énoncé peut consister dans le seul *comment*, et inversement :

Figure II-10 – Généralisation de l'ordre des mots par rapport à la structuration en topic-comment (Givón, 1983 : 20)

COMMENT > COMMENT-TOPIC > TOPIC-COMMENT > TOPIC

4.3.5 Topic de l'énoncé ou du discours ?

Pour conclure cette section sur les définitions de thème/topic, mentionnons Brown & Yule (1988 [1983]), qui réfutent l'idée que *topic* puisse être réductible à un seul constituant d'un énoncé, et qui

s'intéressent davantage au topic discursif, à 'ce dont on parle dans une conversation'. Ils soutiennent que « it is not sentences that have topics, but speakers » (1988 [1983] : 71). Il y a là deux questions distinctes, celle de l'unité linguistique qui exprime ou représente le topic, et celle du domaine d'application du topic. Des approches diverses ont été adoptées par rapport à ces deux questions : ainsi, le topic a été assimilé soit à une expression référentielle, soit à un référent, qui peut ou non être réalisé dans l'énoncé en question (cf. les définitions de Gundel et Lambrecht en section 4.3.3.4 ci-dessus). Dans d'autres approches, comme le rappellent Brown & Yule, *topic* correspond à une proposition, et non pas à un constituant. Dans le premier cas, *topic* est compris comme un référent ou comme l'expression référentielle qui y renvoie, dans le deuxième, *topic* ressemble davantage à la notion de *presupposition* chez Chomsky et Jackendoff par exemple (cf. section 4.3.1.3 ci-dessus). Nous avons déjà évoqué la distinction entre référent topical et l'expression linguistique de celui-ci dans la section précédente ; dans la présente section, nous discuterons, en entrée, la définition de topic comme une proposition. Liée à cette question de l'unité est celle du domaine d'application du topic : l'énoncé ou le discours ? Certains auteurs se sont intéressés exclusivement au topic de l'énoncé (sentence topic), d'autres au topic d'une unité discursive plus large, d'autres encore combinent les deux visées. Dans ce qui suit, nous allons donc présenter une sélection des positions défendues.

L'on peut donc recenser plusieurs approches en ce qui concerne l'unité admise en tant que topic. Selon Lyons (1977b) et Brown & Yule (1988 [1983]), identifier une expression référentielle dans l'énoncé comme topic remonte aux réflexions philosophiques d'Aristote, et à l'envie de découper la phrase pour en distinguer les différentes parties constitutives, comme nous avons pu le mentionner en début de ce chapitre. Lyons rappelle qu'Aristote distingue entre individus et propriétés, et l'analyse de quelques types de phrases isolées induit alors l'idée que le thème/topic correspondrait à un référent explicitement mentionné dans l'énoncé, en tant que support pour la prédication.

L'identification d'une proposition complète qui représenterait le topic se trouve, quant à elle, dans des approches assez différentes : dans des travaux qui opèrent avec le concept de présupposition logique d'une part, et d'autre part dans des travaux davantage praxéologiques, centrés sur les activités sociales accomplies à travers le langage, et qui prennent comme objet d'étude les pratiques et procédures humaines, socialement situées. Or, dès lors que le chercheur s'intéresse aux pratiques langagières, à ce que les locuteurs accomplissent en parlant, il semble cohérent que le topic soit exprimé en termes procéduraux, et non seulement référentiels.

4.3.5.1 La notion de topic du discours chez Ochs Keenan & Schieffelin

L'approche de Ochs Keenan & Schieffelin (1976b) n'est pas facile à classer par rapport à ces considérations. Les auteures utilisent le terme *discourse topic*, qu'elles définissent comme « the proposition· (or set of propositions) about which the speaker is either providing or requesting new

information » (1976b : 338). En même temps, elles précisent que « each declarative or interrogative utterance in a discourse has a specific discourse topic » (1976b : 343). Dès lors, leur proposition ressemble d'assez près à la présupposition d'un énoncé, tel que définie par Chomsky et Jackendoff (cf. section 4.3.1.3 ci-dessus). Les auteures ont d'ailleurs recours au terme *presupposition pool*, à partir duquel les locuteurs choisiraient leur *discourse topic*. En même temps, Ochs Keenan & Schieffelin (1976b) ne se préoccupent pas tellement des présuppositions logiques sous-jacentes à un énoncé, mais davantage à ce que les locuteurs accomplissent en produisant un énoncé. Ainsi, elles notent que « in determining the discourse topic of an utterance, it is useful to determine the purpose or reason behind each utterance », précisant ensuite que la tâche qui incombe à l'interlocuteur serait « to determine what question the speaker may be answering » : un locuteur qui entendrait des voitures s'entrechoquer et qui dirait *An accident* répondrait ainsi, selon les auteures, à la question implicite *What was that noise ?* (1976b : 343). La distinction entre topic de l'énoncé et topic discursif est justifiée, entre autres, par l'exemple suivant :

Exemple II-31 – Topic de l'énoncé vs. topic discursif (tiré de Ochs Keenan & Schieffelin, 1976b : 381)

Locuteur A : What's the matter?
 Locuteur B : My father, he's bugging me again.

Les auteures expliquent alors que le topic discursif pour les deux énoncés correspondrait à la proposition 'something is the matter'. L'expression référentielle disloquée *My father, he*, si elle correspond au topic de l'énoncé du locuteur B, ne constitue pas, en revanche, le topic du discours en train de se tenir entre les locuteurs A et B. Il est intéressant de noter que ce type d'exemple est fréquemment employé pour illustrer des énoncés qui ne comporteraient pas de topic, des énoncés thétiques ou *out of the blue*, répondant précisément à des questions d'ordre très général comme celle du locuteur A (voir aussi la notion de topic cadratif en section 4.3.3.5 ci-dessus). Il devrait alors être surprenant d'y trouver un SN disloqué, réputé comme marqueur de topic. Nous reviendrons à cette question lors de la présentation générale des constructions disloquées dans le CHAPITRE IV, ainsi que lors de la discussion de nos résultats.

4.3.5.2 Coexistence de topic de l'énoncé et topic discursif dans le modèle de Van Dijk

Van Dijk (1977), au contraire de Ochs Keenan & Schieffelin (1976b), qui travaillent avec des données de conversation spontanée, étudie des textes monologiques écrits. Pour van Dijk (1977) également, *discourse topic* n'est pas un constituant, mais une proposition, qui résume les aspects les plus importants du discours en question, et qui pourrait répondre à une question comme « what's the story about? » (1977 : 56). A la différence de l'étude de Ochs Keenan & Schieffelin déjà mentionnée, van Dijk opère avec les deux notions de *sentence topic* et *discourse topic* (parfois *textual topic*) :

« At the level of the sentence, a topic is a specific function assigned to some part of a (possibly compound) proposition and indicates the way information is linearly distributed, whereas a textual topic indicates how information is globally

organized. In the first case, the topic is the link, between given information and new information, for each sentence in the discourse, whereas the textual topic is the hierarchical organization of the whole of information of all sentences, taken at the same time. » (Van Dijk, 1977 : 59)

4.3.5.3 *Topic de l'énoncé, Topic discursif et Objets de Discours*

Il convient enfin de citer l'approche conversationnaliste du topic de Berthoud et Mondada notamment (Berthoud & Mondada, 1993, 1995; Mondada, 1994, 2002; Mondada & Dubois, 1995). Dans cette approche, les auteures se distancient de l'analyse du topic dans le cadre de la phrase ou de l'énoncé, et se situent dans un cadre qu'elles qualifient de discursif, énonciatif et constructiviste (Berthoud & Mondada, 1993, 1995). Discursif, parce qu'elles s'intéressent au développement du topic au cours de l'élaboration du discours ; énonciatif, car elles considèrent non pas le topic en tant que tel, mais le topic tel que l'énonciateur le conçoit : « Les objets de discours sont notamment structurés de manière à porter la trace de la perspective que les énonciateurs exercent sur eux - ainsi que de celle qu'ils présupposent chez leur énonciataire - perspectives constitutives des objets, qui délimitent les topics et les informent » (Berthoud & Mondada, 1995 : 207). Constructiviste enfin, parce que « le topic est conçu non pas comme étant posé une fois pour toutes mais comme étant constamment en élaboration » (Berthoud & Mondada, 1993 : 125), négocié entre les locuteurs. Par rapport aux approches du topic que nous avons pu présenter jusqu'ici, l'approche conversationnaliste ne raisonne pas en termes de référence, de « relation entre les mots et les choses » (Mondada, 2002 : 9), mais en termes de référenciation, c'est-à-dire la « relation intersubjective et sociale au sein de laquelle des versions du monde sont publiquement élaborées, évaluées en termes d'adéquation aux finalités pratiques et aux actions en cours des énonciateurs » (2002 : 9; voir aussi Mondada & Dubois, 1995). Les topics, dans cette approche, ne renvoient donc pas à des expressions référentielles ni à des référents, mais à des *objets de discours*,⁵³ c'est-à-dire

« des entités qui sont interactivement et discursivement produites par les participants au fil de leur énonciation [...] constituées dans et par les formulations discursives des participants [...] qui [...] émergent et s'élaborent progressivement dans la dynamique discursive. Autrement dit, l'objet de discours ne renvoie pas à la verbalisation d'un objet autonome et externe aux pratiques langagières ; il n'est pas un référent qui aurait été codé linguistiquement. » (Mondada, 2002 : 9)

Concernant le lien entre la notion de thème/topic et celle d'objet de discours, Berthoud & Mondada (1993) gardent le terme *thème* pour le topic de l'énoncé. Pour Mondada (2002), le *topic* constitue un cas spécial d'objet de discours, précisément « des objets considérés et manifestés par les participants comme ce à propos de quoi ils énoncent le discours en train de se faire » (Mondada, 2002 : 9). Topic ne serait alors pas une catégorie, pragmatique ou linguistique définissable en tant que telle, mais

⁵³ (cf. Mondada, 1994; cf. aussi Sitri, 2006)

saisissable uniquement en tant que trace des activités des locuteurs, une « entité pertinente pour les locuteurs » (Mondada, 2002 : 9).

Le lien entre objet de discours et thème est aussi développé dans un article de Sitri (2006), qui propose que l'élaboration des objets de discours puisse se faire justement à travers des opérations de thématisation au niveau de l'énoncé. Elle étudie donc des constructions spécialisées de la thématisation, et en particulier ce qu'elle nomme les « constituants thématiques multiples, que l'on trouve par exemple dans un énoncé comme *Mon frigo le congélateur il est en bas* » (2006 : 118). Sitri se positionne alors par rapport aux travaux de Mondada, considérant que, si les deux approches ont des points en commun, « la différence essentielle par rapport à cette position nous semble cependant résider dans la place accordée à la langue comme système » (Sitri, 2006 : 127). Alors que Mondada privilégierait l'étude des activités langagières à l'étude du langage en tant que système, Sitri propose que « c'est dans et par les formes linguistiques que se construisent les objets de discours » (2006 : 127).

Nous avons déjà annoncé dans le CHAPITRE I que nous parlerons de référent plutôt que d'objet de discours, et les référents dans des conversations avec des jeunes enfants sont pour la plupart des objets concrets de l'environnement physique. De plus, ce chapitre a permis de montrer que, s'il n'y a pas corrélation parfaite entre les différents axes de la structuration informationnelle et des formes linguistiques (qu'il s'agisse de la forme de l'expression référentielle ou d'autres facteurs formels comme la réalisation prosodique ou la linéarisation de l'énoncé), il existe néanmoins certaines tendances de recoupement non-arbitraire. Nous allons donc tenir compte des dimensions formelles, informationnelles et interactionnelles en adoptant une définition pragmatique du topic de l'énoncé, mais en tenant compte aussi, dans une analyse croisée, comment les formes employées signalent ce que les locuteurs sont en train de faire. Cette dimension dépasse en effet la seule question de l'identification d'un référent. Nous y viendrons dans le CHAPITRE X.

Enfin, le lien entre topic de l'énoncé et un topic discursif global ou la construction d'un objet de discours global, en suivant les analyses par exemple de van Dijk et de Sitri mentionnées ci-dessus, nous paraît très prometteur aussi pour l'analyse de nos données, mais par manque de place, nous devons laisser la dimension des topics du discours et leur lien avec les topics de l'énoncé à des travaux futurs.

4.4 Structuration de l'information et linéarité : synthèse et élargissement

Il convient ici de faire un point sur la perception des liens entre structuration informationnelle, structuration thématique et linéarité. Nous avons vu dans les différentes approches discutées que souvent l'information donnée et/ou le thème/topic étaient associés avec une position initiale dans l'énoncé. Cette association pouvait être formulée comme règle absolue (pour la position initiale du thème chez Halliday par exemple) ou comme une tendance, observée à travers des langues diverses

(cf. les travaux de Chafe, Gundel, Givón, Kuno par exemple). Givón (1988) relativise les liens entre information et ordre des mots dans une étude cross-linguistique, dont il dégage plusieurs principes généraux :⁵⁴ premièrement, il observe que « More predictable/accessible information will tend to be *post-posed* ('moved to the right'); less predictable/accessible information will tend to be *pre-posed* ('moved to the left') » (1988 : 252; les mises en relief sont de l'auteur). Ce principe du codage de l'accessibilité de l'information sera ensuite combiné avec un principe de l'importance de l'information, tous les deux compris sous une contrainte de l'attention et une contrainte comportementale qui voudrait que l'on s'occupe d'abord de la tâche qui revêt de la plus grande importance. L'auteur aboutit donc à un principe qui stipule qu'une information serait exprimée en position initiale dans un énoncé sous deux conditions possibles : si l'information est moins prévisible et/ou si l'information est plus importante.⁵⁵

Nous avons pu voir dans les sections précédentes qu'une définition du topic de l'énoncé en fonction de sa position initiale dans l'énoncé, ou en fonction de son statut informationnel comme donné ne permet pas toujours de cerner correctement cette notion. Bien sûr, les caractéristiques proposées correspondent à une réalité observable ; il semble donc y avoir une certaine corrélation entre topicalité d'un côté et la position initiale dans l'énoncé, la fonction sujet et un statut informationnel donné ou accessible de l'autre côté. Mais ces traits semblent constituer davantage une tendance qu'une contrainte, et ne peuvent servir de définition. Nous nous inscrivons donc dans le dernier groupe de définitions présenté en section 4.3.3 ci-dessus, qui conçoivent le topic comme l'expression d'une relation pragmatique et définissent le topic en termes de *aboutness* ou 'ce sur quoi porte le propos de l'énoncé'. L'un des objectifs poursuivis dans notre thèse sera alors d'examiner les relations entre la dimension du topic-commentaire, saisi en termes pragmatiques, et les statuts attentionnels des référents, régularités de l'alignement syntaxique et choix des expressions référentielles. Afin de faciliter la lecture et éviter au lecteur des allers-retours entre les chapitres, nous aborderons plus en détail les critères définitoires retenus pour le topic pragmatique de l'énoncé au début du CHAPITRE IX, consacré à l'analyse de l'expression du topic dans nos données.

⁵⁴ Les données analysées ainsi que les analyses rapportées d'autres études dans l'article de Givon (1988) sont issues de « text-based quantified studies » (1988 : 279), sans que la nature exacte de ces données soit spécifiée. La généralisabilité de ces résultats est donc limitée.

⁵⁵ cf. les mesures discursives de topicalité de Givón (1983) exposées en section 4.3.4 ci-dessus. Nous reviendrons à cette question sous un autre angle lorsque nous aborderons le lien entre matériel linguistique et structuration de l'information dans la présentation de l'ordre des mots pour le français et pour l'allemand dans le CHAPITRE III.