

Modélisation de la saillance visuelle et linguistique

- DRAFT -

Frédéric Landragin

Résumé

Est saillant ce qui arrive en premier à l'esprit, ce qui capte l'attention. Saillance visuelle ou linguistique, saillance liée à la forme du message ou à son contenu, saillance à effet immédiat ou continu, saillance préalable ou nouvelle, informative ou rhétorique : autant de dimensions de la saillance qu'il s'agit de clarifier pour en déduire une modélisation. Quels sont les recouvrements et les rapports entre ces dimensions ? Quels aspects de la saillance est-il possible de modéliser pour l'interprétation et la production de messages linguistiques liés à l'environnement visuel ? Nous proposons dans cet article une caractérisation de la notion de saillance fondée sur les dimensions retenues, ces dimensions constituant autant de plans de traitement de la saillance. En nous focalisant sur la dimension physique, nous montrons comment aboutir à une formalisation de la notion de saillance à l'aide de scores numériques, formalisation qui constitue un premier pas vers la réalisation de systèmes de détection automatique des entités saillantes.

Mots Clefs : Caractérisation de la saillance ; P-saillance ; C-saillance ; thème et rhème.

1. Introduction

Un élément visuellement saillant, c'est un élément qui ressort prioritairement lors de la perception visuelle d'une scène, au point de prendre une importance cognitive particulière et d'occulter les autres éléments de la scène. Un élément linguistiquement saillant, c'est un élément qui ressort prioritairement lors de la compréhension d'un énoncé, au point de prendre une importance cognitive particulière et, par exemple, d'axer sur cet élément la réaction ou la réponse à l'énoncé.

Que ce soit lors de la lecture d'une image ou d'un texte, ou plus généralement dans toute situation de communication, la saillance intervient fréquemment. Elle permet par exemple l'utilisation et l'interprétation de pronoms lorsque l'entité dont il est question est saillante à la fois pour le locuteur et l'interlocuteur. L'énoncé « il joue bien » en parlant d'un acteur se comprend ainsi sans ambiguïté lorsqu'un acteur particulier a été rendu saillant par une mise en valeur visuelle ou par un énoncé antérieur. C'est sur ce constat que s'appuient certains travaux linguistiques tels que [2], pour proposer des

modèles d'interprétation des pronoms. Le champ d'application de la saillance est cependant bien plus étendu : la saillance peut intervenir dans toute interprétation et dans toute production linguistique, en tant que point de départ pour la construction du sens (interprétation) et pour l'articulation d'un message (génération).

Il n'existe cependant aucune théorie globale de la saillance, qui tienne compte à la fois des modalités visuelle et linguistique ainsi que des aspects contextuels et cognitifs. En linguistique, les principaux travaux qui s'intéressent à cette notion portent sur la structure informationnelle avec les notions de focus, de thème ou de topique qui recoupent celle de saillance sans s'identifier à elle (cf. par exemple [1], [3] ou [4]). En psychologie de la perception, la saillance est souvent liée à l'attention et n'est pas intégrée aux autres aspects intervenant dans la compréhension du langage (cf. [6] par exemple). C'est à la conception d'une telle théorie qu'il nous semble important d'orienter des recherches. Nous proposons dans cet article d'explorer les différentes dimensions que recouvre la saillance, de manière à en déduire une caractérisation qui puisse aboutir à un modèle de calcul des saillances visuelle et linguistique pour l'interprétation et la production de messages linguistiques liés à l'environnement visuel.

2. Les dimensions de la saillance

Nous avons parlé de **saillance visuelle** et de **saillance linguistique**. La modalité de perception constitue ainsi une première dimension dans notre caractérisation de la saillance. Les facteurs qui concourent à la saillance d'un objet visible ou d'une entité du discours sont en effet spécifiques à la modalité considérée. Parmi les facteurs de saillance visuelle se trouvent ainsi les contrastes de luminosité et de couleur, les caractéristiques physiques particulières de l'objet telle qu'une taille importante, ou encore le fait que l'objet se trouve particulièrement isolé dans la scène. Parmi les facteurs de saillance linguistique se trouvent l'accentuation prosodique, l'utilisation d'un présentatif (« c'est cet acteur qui joue bien »), le placement en début d'énoncé, la fonction grammaticale sujet ou encore le rôle thématique privilégié compte tenu de la structure verbale. Bien que les facteurs visuels et linguistiques puissent se regrouper dans des catégories communes comme le placement à un endroit stratégique [5], la matérialisation et par conséquent le calcul de l'importance de ces facteurs dépendent de la modalité considérée. D'où ce premier trait caractéristique de la saillance.

La saillance peut intervenir au niveau de la forme ou au niveau du contenu. Nous retenons ainsi cette distinction comme deuxième trait caractéristique de la saillance. Au niveau de la forme, ce sont les caractéristiques de l'objet visible (forme, taille, couleur, texture) et les caractéristiques lexicales et syntaxiques de l'énoncé linguistique qui interviennent. Ces caractéristiques ont une trace physique et nous parlons ainsi de **saillance physique** (ou P-saillance, cf. [5]). Au niveau du contenu, ce sont des aspects cognitifs qui entrent en jeu, tels que la dérivation d'inférences et les possibilités d'association de l'objet ou de l'entité du discours à des représentations mentales. Nous parlons alors de **saillance cognitive** (ou C-saillance). Comme nous le verrons plus loin, cette distinction reflète des degrés de difficulté dans l'identification des facteurs de saillance : évaluer la P-saillance se fait sur des indices physiques objectifs, alors qu'évaluer la C-saillance se fait sur des hypothèses sur les états mentaux de l'individu considéré. Dans les systèmes de traitement automatique du langage (TAL), seule la P-

saillance est calculable de manière fiable.

Comme troisième trait caractéristique de la saillance, nous retenons la dimension temporelle, avec une distinction entre saillance à effet immédiat et saillance dont l'effet se construit de manière incrémentale dans la durée. L'étude de l'image fixe et de l'énoncé linguistique relève *a priori* de la **saillance à effet immédiat**, alors que l'image animée et le discours ou la conversation relèvent de la **saillance à effet continu**. Nous limitons ainsi l'effet immédiat à, au mieux, la prise en compte d'une entité privilégiée (thème) et d'un fond ou commentaire (rhème). Cette analyse en termes de thème et de rhème convient à l'énoncé linguistique, mais pas forcément à l'image fixe. L'analogie est pourtant aisée dans le cas de l'image de publicité qui cherche à mettre en valeur le produit à vendre (thème) et à donner des arguments ou commentaires visuels pour favoriser la vente (rhème). Comme l'énoncé linguistique, plus exactement l'unité pragmatique, l'image fixe correspond aussi à un seul acte de communication. Au contraire, dans l'image animée et dans le discours, plusieurs actes de communication concourent à la construction de l'histoire, du sujet, ou encore du propos [1]. L'intervention de la saillance s'avère alors plus complexe, avec par exemple la construction de structures hiérarchiques (macrostructures, superstructures) beaucoup plus difficiles à formaliser.

Une autre distinction fondamentale est celle entre l'exploitation d'une saillance déjà existante et la mise en saillance d'un nouvel élément. Cette distinction entre **saillance préalable** et **saillance nouvelle** traduit en quelque sorte les deux étapes classiques de perception et d'action qui sont mises en jeu alternativement. La saillance préalable, qu'elle soit visuelle ou linguistique, est généralement exploitée de manière implicite comme avec l'utilisation d'un groupe nominal anaphorique (« l'acteur »), qui peut *a priori* référer directement à une entité non saillante (et alors singularisée par le groupe nominal). La saillance apporte dans ce cas un antécédent qui permet de soulever l'hypothèse de l'anaphore. La saillance nouvelle consiste à mettre explicitement en avant, à l'aide par exemple d'un présentatif, une entité qui pourra ensuite être reprise facilement. Elle a donc un rôle de préparation à une action future, par exemple de reprise référentielle.

Cinquième et dernière dimension de notre caractérisation, la saillance qui intervient sur un plan informatif (ou **saillance informative**) et se distingue de la saillance qui intervient sur un plan rhétorique (ou **saillance rhétorique**). Au niveau informatif, un élément nouveau est saillant de par l'information qu'apporte cette nouveauté. C'est le cas d'un objet qui devient visible dans la scène, ou d'une nouvelle entité du discours. Au niveau rhétorique, un élément est saillant non par sa nouveauté mais par la façon dont il est mis ou remis en valeur. Pour un message linguistique, cette saillance est liée à la sélection des places les plus efficaces pour l'entité ciblée et les entités qui lui sont liées, et d'une manière générale à l'organisation du discours. Dans le cas d'une scène visuelle, les aspects de mise en scène s'inscrivent dans le plan rhétorique.

3. Modélisation : comment interviennent les dimensions ?

Du fait de quelques recouvrements, les 5 oppositions binaires précédentes ne conduisent pas à 32 cas de saillance, qui seraient de toute façon difficiles à conceptualiser. Ainsi, du côté de la saillance linguistique, la dimension rhétorique se traduit systématiquement par des constructions telles que les dislocations à gauche qui

sont prises en compte dans la dimension physique. De même, la dimension rhétorique consiste souvent à mettre en valeur une entité particulière pour la rendre saillante, ce qui rejoint la saillance nouvelle.

D'une manière générale, il n'est cependant pas possible de réduire une dimension à une autre, comme nous le montrons avec l'exemple suivant : « Pierre, son vélo est cassé », énoncé que nous pouvons imaginer apparaître dans une discussion portant sur un certain Pierre (qui a déjà été mentionné au cours du dialogue). Concernant la première dimension, cet exemple relève clairement de la saillance linguistique. Le vélo de Pierre n'est pas présent, et Pierre non plus : il a par contre déjà été mentionné linguistiquement. Concernant la deuxième dimension, la notion de P-saillance se traduit par la mise en avant de « Pierre » qui vérifie les critères suivants : placement stratégique en début d'énoncé ; dislocation gauche accentuant sa mise en valeur. Même si le vélo bénéficie de la fonction grammaticale sujet, Pierre garde l'avantage. Dans cet exemple, la notion de C-saillance intervient dans le sens que Pierre est activé dans la mémoire à court terme des interlocuteurs, et donc cognitivement saillant. Concernant la troisième dimension, sans autre précision sur le sujet de la conversation, nous nous plaçons ici dans la saillance à effet immédiat. Il se peut néanmoins que la conversation détaille l'histoire de la vie de Pierre, et dans ce cas toute mention supplémentaire de Pierre a un effet de saillance dans la durée. Concernant la quatrième dimension, nous ne pouvons pas dire qu'une saillance préalable soit exploitée, l'énoncé ne comportant ni pronom ni groupe nominal susceptible de reprendre un antécédent préalablement mentionné. Par contre, une saillance nouvelle intervient avec la mise en thème de Pierre. Ce nouveau thème pourra ainsi être repris ultérieurement par un pronom : si nous imaginons l'énoncé suivant « il était vert », « il » reprend en effet plus facilement Pierre que le vélo... Enfin, concernant la cinquième dimension, la facette informative correspond à l'introduction de l'entité du discours « vélo » dont il n'était pas question auparavant, et surtout du fait que ce vélo a été cassé (rhème de l'énoncé, correspondant à l'information nouvelle). La facette rhétorique se limite ici à la construction de l'énoncé incluant une dislocation à gauche, ce qui était aussi l'un des facteurs de P-saillance (mais un seul seulement, ce qui montre bien que ces dimensions ne sont pas équivalentes).

Il s'avère tout à fait possible de fixer des valeurs dans chacune des dimensions et d'imaginer un exemple correspondant. Ainsi, du côté linguistique, si nous nous limitons à la P-saillance et à un effet continu dans le temps, nous entrons par exemple dans le domaine poétique, avec les rimes dont l'effet se fait sentir petit à petit et qui induisent une certaine saillance. De même, si nous considérons la C-saillance à effet continu, nous retrouvons la notion de topique (ou de propos) telle qu'elle est décrite par exemple dans [1]. Du côté de la perception visuelle, la C-saillance à effet continu correspond au topique ou propos d'un film. Sans entrer dans les détails, l'analyse de l'exemple « attention, il risque de te mordre » (énoncé par un gardien de zoo à un enfant qui tend la main dans une cage enfermant un lion) active les aspects suivants : saillance visuelle en même temps que saillance linguistique, C-saillance, saillance à effet immédiat, saillance préalable et saillance informative.

Avec ces exemples, il nous semble que prendre en compte nos dimensions permettrait de mieux appréhender les notions de focus, thème et topique. Les définitions de ces notions dans les travaux portant sur la structure informationnelle sont en effet multiples. Elles sont surtout bien souvent ambiguës et se recoupent du fait d'un manque de précision dans leurs caractérisations. Nous espérons que nos dimensions sont

susceptibles d'apporter un éclairage nouveau sur ces aspects.

4. Modélisation : quels processus mettre en œuvre ?

Il n'est malheureusement pas évident, et peut-être pas pertinent, de gérer autant d'aspects lors de l'interprétation du moindre énoncé. Si nous nous plaçons dans le domaine du traitement automatique du langage, l'intérêt d'un modèle de la saillance réside dans la possibilité d'en déduire une formalisation, par exemple un algorithme de détection des entités visuelles et linguistiques saillantes, de manière à les privilégier lors de l'interprétation, lors de l'élaboration d'un résumé automatique ou pour toute autre application du TAL. Dans le cadre du dialogue homme-machine, un tel algorithme permettrait non seulement de proposer une solution dans un cas ambigu, mais permettrait également de comprendre pourquoi l'utilisateur du système de dialogue a prononcé tel énoncé, prouvant de ce fait qu'il n'a pas vu tel objet ou qu'il n'a pas fait attention à telle entité du discours : en effet, nous nous limitons parfois aux seules entités saillantes, et, si le système est capable d'en faire autant, il ne nous en comprendra que mieux.

Une méthode simple et efficace consiste à gérer des scores numériques de saillance. Nous avons par exemple énuméré un certain nombre de facteurs de saillance visuelle et de saillance linguistique. En supposant que le système soit capable d'identifier les caractéristiques de la scène visuelle telles que les contrastes de couleur et les groupements d'objets, il est alors possible d'affecter à chacun des objets visibles un score de 0 ou de 1 pour chaque facteur de saillance. Un objet particulièrement grand se verra ainsi attribuer un 1 pour le facteur taille, un objet particulièrement isolé un 1 pour le facteur isolement. En faisant ensuite la somme de ces 1, nous arrivons à quantifier la saillance visuelle de chaque objet. De même pour les entités du discours : en attribuant 1 à l'entité ayant la fonction grammaticale sujet et 0 aux autres, en attribuant 1 à l'entité placée en première position, etc., puis en faisant les sommes, nous obtenons des scores de saillance linguistique.

Comme nous le montrons avec ces exemples, la méthode s'avère aisée pour les facteurs de P-saillance visuelle et de P-saillance linguistique. Si elle est théoriquement possible pour les facteurs de C-saillance, elle s'avère cependant beaucoup plus aléatoire puisque, comme nous l'avons évoqué plus haut, le système de dialogue est alors obligé de faire des hypothèses sur les états mentaux de son utilisateur, en particulier sur le contenu de sa mémoire à court terme. Une solution consiste à spécifier des règles fiables d'inférences de telles hypothèses à partir des traces linguistiques. Par exemple, l'emploi d'un pronom reflète la forte accessibilité de l'entité correspondante dans la mémoire à court terme. Des difficultés similaires se posent pour la formalisation des autres dimensions de la saillance : si la saillance à effet immédiat ne pose pas de problème supplémentaire, la saillance à effet continu nécessite la spécification des facteurs intervenant dans la durée, ainsi qu'une gestion incrémentale de scores pour ces facteurs. Nous retiendrons en particulier la fréquence d'apparition en première position d'une entité dans les différentes phrases d'un discours en tant que facteur de saillance à effet continu. En ce qui concerne la saillance préalable et la saillance nouvelle, il s'agit de distinguer les facteurs privilégiant la première (en particulier la reprise anaphorique) et les facteurs privilégiant la seconde (thématisation, topicalisation). Enfin, en ce qui concerne les aspects informatifs et rhétoriques, il est toujours possible de commencer une

formalisation par la prise en compte des facteurs les plus simples, c'est-à-dire la prise en compte des premières apparitions des entités de discours pour la saillance informative, et la prise en compte d'une bibliothèque de structures de phrases pour la saillance rhétorique.

5. Conclusion et perspectives

Après avoir défini cinq plans d'analyse de la notion de saillance de manière à caractériser cette notion plus précisément que ne le font les travaux portant sur la structure informationnelle, nous avons détaillé une modélisation à base de facteurs de saillance et de scores numériques permettant de quantifier les différentes facettes de cette notion. Cette quantification présente des inconvénients, dans la mesure où certains facteurs ne sont pas identifiables, d'autres interviennent à plusieurs niveaux (nous avons par exemple fait intervenir la reprise pronominale dans plusieurs dimensions) et d'autres s'avèrent trop réducteurs par rapport aux multiples possibilités que présente le langage. Les perspectives de ce travail sont ainsi d'une part la spécification précise des rôles des facteurs de saillance, et d'autre part l'extension de cette formalisation de manière à proposer à terme un outil d'identification automatique des entités saillantes dans l'image ou le texte.

Références

- [1] A. Grobet, *L'identification des topiques dans les dialogues*, Duculot (2002).
- [2] B.J. Grosz, A.K. Joshi et S. Weinstein, Centering: A Framework for Modelling the Local Coherence of Discourse, *Computational Linguistics* 21(2), p. 203-225 (1995).
- [3] E. Hajičová, T. Hoskovec et P. Sgall, Discourse Modelling Based on Hierarchy of Saliency, *Prague Bulletin of Mathematical Linguistics* 64, p. 5-24 (1995).
- [4] K. Lambrecht, *Information Structure and Sentence Form. Topic, Focus and the Mental Representation of Discourse Referents*, Cambridge University Press (1994).
- [5] F. Landragin, Saillance physique et saillance cognitive, *Cognition, Représentation, Langage* 2(2), <http://edel.univ-poitiers.fr/corela> (2004).
- [6] G.A. Rousselet et M. Fabre-Thorpe, Les mécanismes de l'attention visuelle, *Psychologie Française* 48(1), p. 29-44 (2003).
- [7] R.J. Stevenson, The Role of Saliency in the Production of Referring Expressions, In: K. Van Deemter et R. Kibble (eds.) *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, CSLI Publications, p. 167-192 (2002).