

# Enjeux d'un corpus oral comparable français-chinois : l'exemple de *je ne sais pas* / *wǒ bù zhīdào*

**Rou CHEN**

STL, UMR-CNRS-8163, Université de Lille  
rou.chen@univ-lille.fr

This study introduces a comparable corpus of spoken French and spoken Chinese designed to support cross-linguistic research. To illustrate its analytical potential, we compare the use of the expressions *je ne sais pas* in French and its semantic equivalent *wǒ bù zhīdào* in Chinese. Through a semantic, pragmatic, and discursive examination, we explore how these expressions, beyond their basic function of signalling ignorance, are employed in various interactional contexts. The findings highlight both morphosyntactic and pragmatic similarities and divergences between the two languages, which may inform the design of pedagogical activities aimed at enhancing Chinese learners' pragmatic competence in French as a Foreign Language.

## 1. Introduction

Les corpus oraux constituent aujourd'hui des ressources précieuses pour une diversité de domaines en linguistique. Leur exploitation a permis d'approfondir la compréhension des interactions verbales (Traverso, 1996 ; Kerbrat-Orecchioni, 2005), des phénomènes sociolinguistiques tels que la politesse (Terkourafi, 2002 ; Spencer-Oatey & Jiang, 2003), de la structuration du discours oral spontané (Baude et al., 2006) ou encore de la distribution de marqueurs linguistiques peu présents ou utilisés de manière différente dans les corpus écrits (Delahaie, 2009 ; Pekarek Doehler, 2016 ; Tutin, 2019). Ils offrent également un accès privilégié à la langue en usage, facilitant l'analyse grammaticale en contexte (Debaisieux, 2005).

Cette richesse observée dans le domaine de la recherche linguistique a conduit à une valorisation croissante des corpus oraux dans le champ de la didactique des langues. Les travaux du CRAPEL ont montré qu'il était possible de proposer une réelle démarche d'apprentissage du français à partir de l'observation et de l'exploitation directe de corpus oraux authentiques par les apprenants eux-mêmes (Debaisieux, 2009 ; André, 2023). Ces corpus peuvent être mobilisés selon des modalités variées : en version originale, comme

supports illustrant le caractère spontané, interactionnel et contextuel de la communication (Xu & Zhang, 2016); comme ressources permettant de sensibiliser les apprenants à la variation linguistique (Debaisieux, 2009; Hallion, 2020); ou encore comme base pour la conception de dialogues-modèles adaptés aux besoins spécifiques des apprenants et aux objectifs pédagogiques visés (Delahaie, 2013).

Dans cette perspective, le projet DOC (Didactique, Oral, Corpus), initié en 2018 (Delahaie et al., 2018) au sein du laboratoire *Savoirs, Textes, Langage* (STL) à l'Université de Lille, s'inscrit dans une démarche visant à rapprocher données orales authentiques et pratiques d'enseignement<sup>1</sup>. Une des spécificités de ce projet réside dans son approche interlangue : il vise à exploiter les usages discursifs *in situ* dans plusieurs langues. Pour ce faire, une base de données multilingue est en cours de constitution<sup>2</sup>. Elle comporte à l'heure actuelle des situations d'interactions orales comparables en français, chinois, espagnol et italien : échanges dans des commerces, rencontres fortuites ou de la vie quotidienne, expression d'opinion (débat). Il s'agit de situations de communication dans lesquelles les apprenants sont le plus souvent amenés à mobiliser la langue cible, et qui facilitent par conséquent l'intégration de certains acquis langagiers (Gadet & Guérin, 2008). C'est dans ce cadre qu'un corpus comparable d'interactions orales en français et en chinois a été élaboré, type de ressource inédit à ce jour.

Le présent article se propose d'examiner la valeur pragmatique de ce corpus dans l'analyse comparative entre le français et le chinois. Cette étude constitue une étape préalable à son exploitation didactique dans l'enseignement du FLE aux apprenants sinophones, en vue de développer leur compétence pragmatique. Dans cette optique, la première partie est consacrée à la méthodologie de constitution du corpus et à la présentation des

---

<sup>1</sup> <https://stl.univ-lille.fr/productions-scientifiques/ressources>

<sup>2</sup> Une partie des données est d'ores et déjà accessible sur la plateforme *Ortolang* (<https://www.ortolang.fr/market/corpora/doc-stl>), et plusieurs travaux portent sur leur exploitation, notamment une étude consacrée aux verbes faibles en français et en espagnol (Delahaie, Rouanne, et Canut 2022).

données. La seconde propose une analyse sémantico-pragmatique et discursive des expressions quasi équivalentes, à savoir *je ne sais pas* en français et *wǒ bù zhīdào* en chinois, analyse qui permettra de dégager des pistes didactiques abordées dans la conclusion.

## 2. Présentation du corpus

Un corpus comparable du français et du chinois parlés a été constitué entre 2019 et 2021 (Chen, 2024). Ce corpus comprend 72 conversations enregistrées en langue première de binômes français et de binômes chinois, âgés de 18 à 31 ans<sup>3</sup>. Il totalise environ 170 minutes en français (33 169 mots) et 146 minutes en chinois (36 190 mots)<sup>4</sup>. Ces conversations ont été transcrites à l'aide du logiciel ELAN et sont actuellement en cours d'anonymisation en vue de leur diffusion prochaine sur la plateforme *Ortolang*.

La comparabilité des données entre le français et le chinois est assurée par l'homogénéité des scénarios de conversation, qui comprennent trois situations de communication dans lesquelles les participants sont invités à exprimer leurs opinions, en particulier leurs désaccords : le choix d'un film, la justification de ce choix, un débat sur le thème central du film choisi. Plus précisément, chaque participant sélectionne tout d'abord un film parmi une liste de films illustrés par des affiches, couvrant un large éventail de thématiques (amour, éducation, relations familiales, intelligence artificielle, etc.), avant de négocier avec son binôme celui qui sera visionné ; ensuite, après la projection de la bande-annonce du film retenu, les deux participants partagent leurs impressions sur ce film ; enfin, ils prennent part à un débat portant sur une question liée au film. Par exemple, après la projection de *La Land*, qui traite de l'amour et des rêves, le débat porte sur la question

---

<sup>3</sup>Les participants au sein de chaque binôme se connaissent dans la vie réelle, en tant que camarades ou amis. Les conversations enregistrées dans le cadre de l'expérimentation reposent sur la construction et le maintien d'une relation positive entre eux.

<sup>4</sup>En raison de l'épidémie, certaines conversations sont recueillies par les biais de *Zoom* ou *WeChat*.

suivante : « L'amour peut-il être écrasé par la réalité ? ». Pendant l'expérimentation, les participants ne sont pas informés de l'objectif réel de l'étude, et l'intervention de l'observatrice reste minimale.

La constitution de ce corpus ne se limite pas à une simple collecte de données linguistiques : elle vise également à fournir des données pragmatiques, c'est-à-dire des éléments liés à l'usage concret de la langue en interaction, tels que les actes de parole, les marqueurs discursifs ou les stratégies de (dés)accord. Ces données sont importantes pour développer la compétence pragmatique des apprenants, entendue comme la capacité à mobiliser les actes de langue de manière appropriée dans un contexte socioculturel donné, à reconnaître les écarts de valeurs entre interlocuteurs, et à les négocier pour aboutir à une compréhension mutuelle (Nguyen, 2011). Comparer des expressions présentant des proximités sémantiques mais des différences d'usage interactionnel offre une voie privilégiée pour exploiter ce corpus et pour révéler les traits communs et distinctifs du fonctionnement pragmatique du français et du chinois. À titre d'exemple, nous analyserons deux séquences qui indiquent l'ignorance dans chacune des deux langues : *je ne sais pas* (désormais JSP) et *wǒ bù zhīdào* (désormais WBZD)<sup>5</sup>. Bien qu'elles partagent un sens littéral similaire et soient souvent utilisées pour atténuer des opinions, leurs usages ne sont ni systématiquement équivalents, ni culturellement interchangeables.

### 3. ***Je ne sais pas* et *wǒ bù zhīdào* : similaires mais non équivalentes**

Nous présentons dans ce qui suit quelques données quantifiées qui rendent compte de l'intérêt de comparer JSP avec WBZD, et nous portons ensuite notre attention sur leur emploi factif et non factif, y compris leur emploi en tant que marqueur discursif.

---

<sup>5</sup> JSP désigne l'ensemble des réalisations morphophonologiques de la séquence *je ne sais pas*, y compris les formes réduites telles que l'omission de la négation *ne* (*je sais pas*). De même, WBZD regroupe toutes les variantes morphophonologiques de la séquence *wǒ bù zhīdào*, notamment celles comportant l'omission du pronom personnel *wǒ* ou du verbe *dào* après *zhī*.

### 3.1 Quelques données quantifiées

Nous avons recensé 78 occurrences de JSP et 33 occurrences de WBZD considérées comme des formes unitaires (sans insertion de qualificatifs entre les constituants de la séquence). Après élimination des énoncés tronqués, nous retenons respectivement 72 occurrences de JSP et 28 occurrences de WBZD pour l'analyse syntaxique. Leur répartition syntaxique est présentée dans le tableau suivant :

Configuration syntaxique <sup>6</sup>	JSP	WBZD
x p/p x	30 (42,9 %)	5 (17,9 %)
x en incise	14 (19,4 %)	0 (0,0 %)
x si p	13 (18,1 %)	11 (39,3 %)
x ce que p	5 (6,9 %)	3 (10,7 %)
x pourquoi/comment p	4 (5,6 %)	1 (3,6 %)
x seule en réponse à l'assertion/la question fermée	3 (4,2 %)	5 (17,9 %)
x question directe	2 (2,8 %)	0 (0,0 %)
x o/o x	1 (1,4 %)	3 (10,7 %)
Total	72 (100 %)	28 (100 %)

Tab. 1 : Configurations syntaxiques de JSP et WBZD

Les données révèlent que la configuration antéposée (1) ou postposée (2) par rapport à la proposition *p* est la plus fréquente pour JSP (42,9 %), suivie par l'emploi en incise (19,4 %) (3), et l'introduction d'une subordonnée avec *si* (18,1 %).

- (1) A **je sais pas** je préfère largement ce film (DOC \_2020\_FC02)<sup>7</sup>
- (2) A je choisis un exemple très très enfin très violent enfin très très catégorique mais il y a plein d'autres choses enfin **je sais pas** (DOC \_2020\_FT02)
- (3) A on a du mal à intégrer tout le monde euh : : *je sais pas* et : : à vivre ensemble **je sais pas** si tu as vu : : Unbelievable (DOC \_2020\_FC04)

En revanche, WBZD est le plus souvent utilisé dans des constructions de type *x si p* (39,3 %), comme celle en (4), suivi par son emploi seul en réponse à

<sup>6</sup> La configuration x o/o x renvoie aux usages de l'expression avec COD, tels que *je ne sais pas son nom* et son équivalent en chinois.

<sup>7</sup> La convention de transcription est appliquée dans l'ensemble des conversations : « [...] » signale les chevauchements, le signe « : » marque l'allongement vocalique, « &...& » encadre un segment produit en riant, et « SD » indique la durée d'un silence, « \ » signifie une petite pause dans le tour de parole, ↑ signifie une voix montante.

une question ou à une assertion (17,9 %), comme en (5), et par son emploi antéposé/postposé à la proposition *p* (17,9 %), comme en (6).

- (4) A **wǒ bù zhīdào** shì bùshì zhège (DOC\_2020\_CC12)  
'*je ne sais pas si c'est cela*'
- (5) A hǎo nà wǒ wèn nǐ : nǐ duì zhège lèixíng de diànyǐng gǎn xìngqù ma?  
'*bon alors je te demande : est-ce que ce genre de film t'intéresse ?*'
- B **wǒ bù zhīdào** (DOC \_2020\_CC01)  
'*je sais pas*'
- (6) B ò shì {xiào} & shì háizi de fùqīn & ò **wǒ bù zhīdào** (DOC\_2020\_CD11)  
'*ah c'est {rire} & c'est le père de l'enfant & ah je ne sais pas*'

Les données reflètent également une diversité syntaxique nettement plus grande pour JSP que pour WBZD. D'abord, JSP apparaît fréquemment en incise à l'intérieur de *p*, jouant un rôle de parenthèse discursive, comme en (3). Ce type d'insertion est absent en chinois oral, où l'interruption de la proposition *p* par WBZD est peu naturelle, comme le montre l'exemple inventé (7) :

(7) ?? wǒmen kěyǐ jiāo xiǎohái, wǒ bù zhīdào, zěnmē zhòng shù.

1PL pouvoir enseigner enfant 1SG NEG savoir comment planter arbre

Par ailleurs, JSP peut précéder une question directe pour marquer une hésitation, comme en (8), alors que ce type d'usage n'a pas d'équivalent naturel en chinois, où la formulation des interrogatives directes et indirectes ne présente pas de distinction morphosyntaxique claire.

- (8) A & on est dans deux extrêmes tu [vois] &  
B [c'est] ça c'est ça mais  
euh : mais je trouve que la France a même une tendance à beaucoup rire des clichés comme ça et euh : : **je sais pas**  
au final ça fait du bien ou du mal  
SD 0.132  
A je pense que ça fait du bien (DOC\_2020\_FC04)

Bien que fondées sur un corpus restreint, les différences de fréquence relevées suggèrent l'existence de divergences dans les fonctions discursives et pragmatiques des deux expressions, ce qui nous motive à comparer leurs

emplois en contexte.

### 3.2 *Fonction factive et non factive*

Les verbes *savoir* et *zhīdào* présentent une factivité dans les contextes affirmatifs : ils présupposent la vérité de la proposition subordonnée (Kiparsky & Kiparsky, 1970 ; Li, 2015 ; Dostie, 2016). En (9) et en (10), le locuteur affirme la validité de l'information évoquée.

(9) Je sais que t'es là.

(10) wǒ zhīdào nǐ zài nǎlǐ  
'je sais que t'es là '

Dans des contextes négatifs, la factivité peut persister, notamment lorsque *savoir* est conjugué à un autre temps que le présent de l'indicatif, comme l'imparfait en (11), ou bien dans certaines interrogatives indirectes où JSP et WBZD apparaissent dans des constructions de type *x ce que p*, *x pourquoi/comment p*, et *x o/o x* comme en (12). Dans ces emplois, JSP et WBZD expriment souvent une ignorance portant sur un événement ou une partie informationnelle de celui-ci.

- (11) A ok↓ ben : moi *je sais pas ce que* t'en penses\mais **je savais même pas** qu'il y avait un film Black-Mirror↑ pour le coup  
B il y a (Ya)↑ un film Black-Mirror ?  
A [ben] apparemment↑ {rire}(DOC \_2020\_FC08)
- (12) B [ouais↓]\c'est pour ça aussi  
A =mais c'est bizarre que j'en aie pas entendu parler [du coup ]  
B [tu penses] ça\ça a l'air léger ben  
oui↓ oui ↓\{rire} [aie pas peur ]  
A [& **je ne sais pas pourquoi**]  
j'en ai pas entendu parler (DOC\_2020\_FC12)

Néanmoins, une différence majeure réside dans le traitement temporel de la factivité. Le français encode la temporalité à travers la flexion verbale (présent, imparfait, passé composé), permettant de distinguer entre ignorance actuelle et ignorance passée. À l'inverse, en chinois, l'ancrage temporel repose sur le contexte discursif, la morphologie verbale étant invariable, comme en (13) :

- (13) B zhège diànyǐng jiùshì gēnjù nàgè gùshì  
gǎibiān de [{xiao :}]  
'ce film est justement adapté de cette histoire  
[{{rire : }}]'
- 2 A [&èrshísì gè bǐlǐ&]  
'[& Les Mille et une vies de Billy Milligan &]'
- 3 B suǒyǐ nǐ qíshí shì bù zhīdào zhège shìqing  
'donc en fait tu ne savais pas ça'  
=wǒ bù zhī\& wǒ zhēnde ↑bù zhīdào & zhège
- 4 A je NEG savoir je vraiment NEG savoir  
ce.CL  
diànyǐng shì èrshísì gè Bǐlǐ gǎi [ biān de]  
film être vingt CL Billy adapter PART.REL  
'=**je savais pas**\& je savais vraiment ↑pas & que ce  
film était adapté de Les Mille et une vies de Billy  
Milligan [en fait]'

(DOC\_2020\_CD10)

Dans cet extrait, A et B ont visionné la bande-annonce du film *Split* et commencent à partager leurs impressions. Lorsque A évoque une proximité thématique entre *Split* et *Les Mille et une vies de Billy Milligan*, B confirme qu'il s'agit bien d'une adaptation et suppose que A n'était pas au courant. A réagit immédiatement en affirmant qu'elle ne le savait pas (ligne 4). On observe ici une ignorance passée de A, puisqu'elle possède désormais la connaissance de cette adaptation. Cette ignorance passée n'est inférée qu'à partir du contexte interactionnel, c'est-à-dire de la manière dont A et B échangent et se répondent dans la conversation.

Enfin, la factivité tend à s'effacer dans certains contextes négatifs. Dans ce cas, JSP et WBZD peuvent suspendre un engagement énonciatif en introduisant une incertitude ou un doute, soit sur sa propre affirmation (14), soit sur celle de l'interlocuteur (15), permettant ainsi de moduler le désaccord de manière plus ou moins implicite. En (14), par l'usage de JSP, B introduit un film en exprimant un doute sur la pertinence de ce qu'il s'apprête à dire par rapport au thème de la discussion. Cette stratégie peut être considérée comme une « stratégie préemptive » (*preemptive strategy*) selon Yin & Yin (2017) : une approche proactive visant à anticiper une objection ou une opposition avant qu'elle ne se produise.

- (14) B après il y a(Ya) un film que je pense  
beaucoup **je ne sais pas si je divague un peu**



- du sujet** mais c'est Qu'est-ce qu'on a fait au Bon Dieu  
SD 0.004
- A ouais [mais j'ai vu aussi sur]
- B [c'est sur c'est sur] des parents racistes qui ont trois filles qui vont toutes se marier avec [un noir un Arabe un juif non quatre filles] (DOC\_2020\_FC4)
- (15) B = nà nán nán zhǔréngōng shì háizi de fùqīn ma ?  
'alors ce ce personnage masculin est le père de l'enfant ?'  
SD 0.424
- A kěndìng shì ya: ↓  
'bien sûr que oui : ↓'  
SD 0.284
- B ò shì {xiào} & shì háizi de fùqīn & ò **wǒ bù zhīdào**  
'ah c'est {rire} & c'est le père de l'enfant & ah **je ne sais pas**'  
SD 0.241
- A bú shì háizi de fùqīn : \nà [nà tā\ yǐ\ tā yǐ]  
'ce n'est pas le père de l'enfant : \ [alors il\ avec \il avec]'
- B [kěnéng shì tā wàigōng \yéye]  
'[peut-être son grand-père maternel\grand-père paternel]'
- A tā yǐ shénme shēnfèn chūxiàn zài zhège yùgàopiàn lǐmiàn (DOC\_2020\_CD11)  
'avec quelle identité il apparaît dans cette bande-annonce'

En (15), WBZD est mobilisé dans un tout autre contexte : B semble hésiter à contredire explicitement l'affirmation de son interlocuteur, selon qui le personnage principal est le père de l'enfant. Le doute exprimé par B à travers WBZD joue un rôle atténuateur : il permet de manifester un désaccord sans le formuler de manière frontale. Ce type de désaccord indirect est d'ailleurs immédiatement interprété comme tel par A, qui relance en demandant avec quelle identité le personnage apparaît dans la bande-annonce. L'emploi considéré s'inscrit dans une stratégie de politesse. Contester directement une assertion revient en effet à remettre en cause la pertinence du propos de l'interlocuteur, ce qui menace la face positive – l'image valorisante que chacun cherche à projeter dans l'interaction (Brown & Levinson, 1987). En évitant une

formulation explicite du désaccord, le locuteur préserve la face de l'interlocuteur et maintient l'harmonie interactionnelle.

Parfois, face à une assertion et à une question fermée de l'interlocuteur, le locuteur peut, en exprimant son incertitude sur la proposition, éviter de donner immédiatement son propre avis, comme WBZD dans (16). Dans l'emploi considéré, A feint de ne pas connaître la réponse, alors qu'il est en mesure de répondre de manière affirmative. Cette stratégie lui permet d'éviter de contredire les préférences exprimées par B (qui souhaite voir un film d'amour), en atténuant ainsi une possible dissonance. Elle contribue également à protéger sa propre face positive (ne pas être contredit).

- (16) B è : mǔzǐ de shì nàzhǒng shì ma  
           'euh : c'est un film mère-fils de ce genre-là ? '  
 A **bù zhīdào** \ nà nǐ de àiqíng jiùshì qīnqíng yě kěyǐ ba  
           '**je ne sais pas**\donc ton amour c'est juste de  
           l'affection familiale ça marche aussi hein ?'  
 B è : kěyǐ a kěyǐ nà jiùshì ma- mother will protect you  
       (10nglaise) zhège [diànyǐng]  
       'euh : c'est possible c'est possible donc alors ce  
       film parle mo mother will protect you (anglais)'  
 A [ò ò shì] kàn zhè hǎibào shì [shì guānyú mǎma  
       de] (DOC\_2020\_FC7)  
       '[ah ah oui] c'est un film sur la mère en regardant  
       cette affiche'

Les constructions *x si p*, *x p/p x* et *x seule* en réponse à l'assertion/la question fermée permettent l'expression de l'incertitude ou du doute/désaccord, mais les séquences JSP et WBZD y occupent des statuts différents. Dans la construction *x si p*, ces séquences peuvent moduler l'acte illocutoire auquel elles sont rattachées, en le transformant en ce que Kerbrat-Orecchioni (2001) appelle « semi-question » ou « semi-assertion » : un acte intermédiaire, situé entre l'assertion affirmative et la question explicite. Cet acte réalise une double fonction : d'une part, il traduit l'incertitude du locuteur face à l'énoncé, d'autre part, il offre un moyen discret de vérifier l'exactitude de ses dires, sans imposer explicitement une réponse. Ce caractère faiblement contraignant tient au fait que l'interlocuteur peut choisir de ne pas répondre à une semi-question sans que cela soit perçu comme un manquement interactionnel, contrairement à ce qu'impliquerait une vraie question, plus directement engageante sur le

plan de la politesse. L'exemple (17) illustre un usage de ce type de WBZD.

- (17) B **wǒ bù zhīdào nǐ yǒu méiyǒu** jiùshì shuō \ zhēnzhèng  
 zhàn zài fùmǔ de zhè ge lìchǎng shàng \ ránhòu qù  
 kǎolù tāmen \ shì zěnmē duìdài nǐ de \ nǐ gāngcái yě  
 shuō guò nàme jiùshì shuō \ nǐ de jiāzhǎng \ duì nǐ  
 \ shì fēicháng yángé de [ nàme duì nǐ de ]  
*'je ne sais pas si tu as enfin\si tu t'es vraiment  
 placée à la position de tes parents\puis as réfléchi  
 à la manière dont ils te traitent\tu l'as aussi dit  
 tout à l'heure\c'est-à-dire\tes parents\envers  
 toi\sont très stricts [et envers toi...]'*
- A [eng]  
 'hm'
- B xīnlíng nàme shòudào le shānghài  
*'ton cœur ton esprit alors a été blessé'*  
 (DOC\_2020\_CT5)

L'extrait (17) s'inscrit dans un débat entre deux sinophones autour de la question fermée « L'adage qui aime bien, châtie bien est-il vrai ou faux ? ». B défend cette idée, tandis que A s'y oppose en s'appuyant sur son expérience personnelle : elle évoque les punitions sévères infligées par ses parents et la blessure affective qu'elles ont provoquée. Dans ce passage, B cherche à faire comprendre à A que le châtement parental peut avoir une justification. En disant « je ne sais pas si tu t'es vraiment placée du point de vue de tes parents puis as réfléchi à la manière dont ils te traitent », B formule une semi-assertion/semi-question : sous couvert d'incertitude et de doute, il sous-entend que A n'a pas envisagé la situation du point de vue de ses parents. L'usage de WBZD introduit ainsi une incertitude feinte, qui adoucit le désaccord tout en préservant la face de A. Le fait que A ne réponde pas et laisse B poursuivre, confirme le caractère faiblement contraignant de cette semi-assertion/semi-question.

Dans les autres constructions, JSP et WBZD peuvent être considérés comme un marqueur discursif, point que nous détaillerons dans la partie qui suit.

### 3.3 Usage en tant que marqueur discursif

Les marqueurs discursifs sont couramment reconnus par leur nature généralement optionnelle sur les plans syntaxique et sémantique. Autrement

dit, leur suppression ne rend pas l'énoncé agrammatical puisqu'ils ne contribuent pas au contenu propositionnel des énoncés auxquels ils sont joints. En (18), si JSP est supprimé, l'énoncé est tout à fait recevable syntaxiquement sans changement sémantique.

- (18) B    j'adore la comédie musicale [c'est]  
         A    [{rire}] je sais [Se] pas moi j'ai vu la bande-  
               annonce de La La Land [anglais] justement et je vois  
               de la danse de la pluie des paillettes et euh et le  
               fait que c'est gnangnan dans ce cas (DOC\_2020\_FC04)

Selon Dostie (2004), les marqueurs discursifs peuvent être classés en plusieurs catégories, dont deux nous concernent, à savoir *marqueur de réalisation d'un acte illocutoire*, et *marqueur illocutoire d'interprétation*. La première désigne des unités pragmatiques qui constituent, en soi, un énoncé et peuvent généralement tenir lieu d'intervention. La deuxième désigne des unités pragmatiques qui guident la lecture ou l'interprétation de l'énoncé, et qui peuvent être antéposées, postposées ou intercalées par rapport à l'énoncé principal.

Dans les deux langues, les séquences JSP et WBZD peuvent être considérées comme *marqueurs de réalisation d'un acte illocutoire* lorsqu'elles sont utilisées seules en réponse à une question fermée ou à une assertion, pour exprimer un doute, un désaccord indirect, voire un évitement face à la proposition *p* de l'interlocuteur. Ce type d'emploi, comme nous l'avons déjà souligné, est étroitement lié aux stratégies de politesse. Il permet de préserver la face de l'interlocuteur tout en marquant une forme de distance ou de désengagement discursif. Il s'inscrit ainsi dans une logique d'atténuation interactionnelle, où le locuteur minimise son engagement tout en maintenant la fluidité de l'échange.

Toutefois, une différence apparaît dans l'extension pragmatique de ces séquences en tant que *marqueurs illocutoires d'interprétation* en association avec la proposition *p* (à savoir les constructions *x p/p x*, et *x* en incise). JSP peut se combiner de manière fluide avec diverses stratégies discursives d'atténuation : marquer une hypothèse, exprimer une subjectivité ou

atténuer/introduire un désaccord. L'exemple (19) illustre comment JSP en antéposition prépare un commentaire critique en tout modulant son intensité, alors que l'exemple (20) montre comment JSP en postposition marque une réponse hypothétique.

- (19) B \ben en fait : : l'histoire je trouve qu'elle a ni queue ni tête je l'ai déjà vu\[elle a ni queue ni tête et puis : euh]  
 A [ouais ouais moi aussi\bah c'est une comédie musicale]  
 B et puis **je sais pas** je trouve que c'est gnangnan ↓  
 (DOC\_2020\_FC11)
- (20) A je sais pas qui c'est qu'il est écrit on s'en fout  
 B Spike Jonze Johansson **je sais [Se] pas** bref  
 (DOC\_2020\_FC10)

En (19), B est en train de formuler une évaluation négative du film *La La Land*. Par l'usage de JSP, il cherche à créer un espace discursif dans lequel la proposition qui suit soit relève de sa subjectivité, soit demeure ouverte à la discussion, en raison des limites de son propre savoir. On pourrait remplacer JSP par des atténuateurs épistémiques tels que *je pense*, *personnellement*, *pour moi*, etc. Cette stratégie d'atténuation permet à la fois de ménager la face de A – en évitant d'imposer un jugement catégorique – et de préserver sa propre face, dans la mesure où la véracité de l'énoncé pourrait être remise en question. En ce sens, l'emploi de JSP permet au locuteur de se décharger de la responsabilité de ses propos, tout en lui évitant l'effort de rechercher un mot ou une formulation plus précise (Aijmer, 1986).

En (20), B adopte une posture collaborative. Face à une semi-question de A, elle propose une réponse. L'ajout de JSP indique qu'elle n'est pas entièrement certaine de l'exactitude de son énoncé. Il s'agit ici d'un emploi elliptique de la construction complète, *je ne sais pas si p* (par exemple : *je ne sais pas si c'est correct*), qui sert à exprimer une incertitude. Cet usage peut être remplacé par *je ne suis pas sûr(e)*, mais non par *selon moi*, ce dernier relevant une prise de position subjective sur la validité d'une proposition, et non une incertitude quant à son exactitude.

La grande flexibilité syntaxique de JSP permet également d'étendre l'emploi à d'autres sens discursifs, comme le rôle de *filler*, sans véritable lien sémantique avec l'expression d'ignorance et d'incertitude — phénomène également mis en évidence dans les travaux de Pekarek Doehler (2016) et Debras (2021). Dans (21), par exemple, JSP intervient en cours d'énoncé, non pour marquer une réelle incertitude, mais pour signaler une hésitation ou un effort de formulation, tout en maintenant le tour de parole. Ce type d'usage, en tant que *filler*, n'est pas naturel et plutôt rare avec WBZD, comme l'illustre l'exemple (7) que nous présentons également en (22).

- (21) B on apprend plus sur : : le passé ça c'est toujours des jolies histoires ça n'en fait pas trop peur : : comment **je sais pas** faire pousser des plantes comment euh il y a de tonnes de trucs à leur dire sans : : qui les terrifient pas et qui euh : : (DOC\_2020\_FT2)

- (22) ?? wǒmen kěyǐ jiāo xiǎohái, wǒ bù zhīdào, zěnmē zhòng shù.

1PL pouvoir enseigner enfant 1SG NEG savoir comment planter arbre

Ce type d'usage pour WBZD est absent des travaux fondés sur des corpus à grands volumes de données (plus de 100 000 mots) comme ceux de Tao (2003), Li (2015) et Yin & Yin (2017). Il semble néanmoins apparaître dans notre corpus chez une locutrice plurilingue fortement exposée au français oral. Dans (23), B choisit le film *Her* et, après avoir indiqué son choix, utilise WBZD comme *filler* pour introduire une justification. L'énoncé qui suit laisse entendre qu'elle doute du fait que ce soit bien ce film qui ait été recommandé par son professeur, ce qui peut nous conduire à interpréter cette occurrence de WBZD comme une bribe elliptique de la seconde séquence *je ne sais pas si c'est cela*. S'il est possible de considérer la première occurrence de WBZD comme un *filler*, nous ne pouvons pas exclure la possibilité d'une influence interlinguistique.

- (23) B wǒ yào kàn zhè ge Her (anglais)\tā ↑SD 0.426 **wǒ bù zhīdào** yīnwèi\qián liǎng tiān shàngkè de shīhou wǒmen 1 āoshī tuījiàn le yī ge diànyǐng jiào Her (anglais) **wǒ bù zhīdào shì bù shì** zhège\[dàn zhège xiě le]  
 'je veux regarder Her (anglais)\elle ↑SD 0.426 **je ne sais pas** parce que\avant deux jours notre professeur a recommandé en cours un film qui s'appelle Her **je**

*ne sais pas si c'est cela* \ [mais il a écrit]'

A [èn :]  
[hm :]

B love \ story (anglais ) [ ránhòu hái xiě le ào]  
'love \ story (anglais ) [ il écrit aussi Oscar]'

A [ò : ]  
[oh : ]

B sīkǎ wǒ juéde yīnggāi hái kěyǐ (DOC\_2020\_CC12)  
'je pense que ce film semble bon'

Pour terminer, on peut avancer que l'extension discursive de JSP pourrait être en partie facilitée par certaines propriétés sémantiques du verbe *savoir* centré sur le cours ou l'aboutissement d'un processus cognitif (Rémi-Giraud, 1986, cité dans Dostie, 2016). En fait, *savoir* peut renvoyer aussi bien à un savoir déclaratif qu'à un savoir procédural. Le savoir déclaratif (ou « savoir que »), désigne des connaissances verbalisables sur des faits ou des événements (Ten Berge & Van Hezewijk, 1999, p. 67). Dans cette veine, il pourrait être mobilisé pour affirmer, infirmer ou commenter une proposition, ce qui facilite l'emploi discursif de *je ne sais pas* pour marquer une attitude cognitive du locuteur face à une information (incertitude, subjectivité, voire un rejet). À l'inverse, le savoir procédural (ou « savoir comment ») concerne des aptitudes incorporées et souvent difficiles à expliciter, telles que « savoir nager » ou « savoir jouer aux échecs » (Ten Berge & Van Hezewijk, 1999, p. 67). Dans ce cadre, en énonçant (24), le locuteur montre souvent sa capacité intégrée à nager, non pas une simple connaissance théorique sur la natation.

(24) Je sais nager.

Cet emploi procédural faciliterait l'évolution de *je ne sais pas* vers son emploi discursif pour organiser les discours, y compris initier le tour de parole en (18), clôturer le tour de parole en (20), ou servir de *filler* en (21).

En revanche, le verbe *zhīdào* ne renvoie qu'au savoir déclaratif. Sa mobilisation pour désigner une compétence pratique produit un effet peu naturel :

(25) wǒ zhīdào yóuyǒng  
IPG savoir nager

*'je connais nager'*

De ce fait, si WBZD peut exprimer une posture épistémique vis-à-vis d'une information, son potentiel d'extension vers des fonctions discursives organisationnelles semble plus limité par rapport à JSP.

Cela dit, les propriétés sémantiques du verbe de base ne sont pas les seuls facteurs susceptibles d'influencer l'évolution discursive de ces deux séquences. L'évolution pragmatique peut s'écarter du sens lexical de départ : de nouvelles fonctions discursives peuvent apparaître indépendamment de la sémantique initiale (Dostie, 2004). Par exemple, en anglais, le verbe *know*, tout comme *zhīdào*, réfère principalement au savoir déclaratif. C'est pourquoi on dit « I can swim » plutôt que « I know swimming ». Pourtant, la séquence *I don't know* peut être utilisée non seulement pour atténuer une assertion ou un désaccord (Tsui, 1991 ; Scheibman, 2000 ; Pichler, 2007 ; Kärkkäinen, 2010 ; Weatherall, 2011 ; Nguyen & Tree, 2025), mais aussi pour structurer les discours en tant que « dispositif de gestion du tour de parole » (*turn-exchange device*) (Pichler, 2007, p. 11), servant à initier, clôturer et maintenir le tour de parole en tant que *filler*, comme en (26).

(26) we got home like really really late like at at like **I don't know** like 2:30 or 3:00  
(Nguyen et Tree 2025 : 72)

Ainsi, si la sémantique du verbe de base peut contribuer à expliquer l'évolution discursive et pragmatique de ces séquences, elle ne suffit pas à elle seule. D'autres facteurs, tels que la relation entre les locuteurs, le type d'interaction, ou le contact des langues que nous venons d'illustrer dans l'exemple (23), méritent d'être examinés dans les futures recherches afin de mieux confirmer et comprendre les différences observées entre JSP et WBZD.

#### 4. Conclusion

À partir d'une analyse contrastive entre *je ne sais pas* et *wǒ bù zhīdào*, cette étude montre l'intérêt de constituer un corpus comparable d'interactions orales en français et en chinois. Leurs usages effectifs en interaction orale vont bien



au-delà de la simple déclaration d'ignorance ou de la négation catégorique d'un savoir ou d'une connaissance. Elles apparaissent comme de véritables ressources discursives, mobilisées pour exprimer l'incertitude, atténuer l'engagement énonciatif, organiser la prise de parole, réaliser un évitement, ou encore gérer poliment et subtilement un désaccord. La comparaison entre les deux séquences a également révélé des écarts structurels, à la fois au niveau de l'expression de la temporalité (notamment en raison de l'absence fréquente de marquage morphologique du temps en chinois), de la flexibilité syntaxique (plus marquée pour JSP que pour WBZD), ainsi que des fonctions discursives (rôle de *filler* pour JSP). Ces écarts, susceptibles d'engendrer des difficultés pour les apprenants chinois, soulignent la nécessité d'une approche didactique attentive à la fois aux spécificités de la langue cible et à celles de la langue maternelle des apprenants.

La poursuite de la recherche vise à exploiter le corpus constitué comme support pour l'enseignement du FLE, conformément à l'objectif principal du projet DOC. Des activités pédagogiques pourraient être conçues, inspirées des résultats de cette recherche, afin de développer la compétence pragmatique des apprenants sinophones, notamment dans l'expression et la gestion des opinions en interaction. Concrètement, il s'agirait de i) familiariser les apprenants avec les différences sémantico-pragmatiques entre JSP et WBZD, ii) identifier leurs effets sur les relations interpersonnelles (maintenir/reconstruire/nuire à une bonne relation) dans des situations concrètes. Ces activités permettraient de mettre en relation les choix linguistiques avec les paramètres contextuels et les objectifs communicationnels, et de construire un système d'interprétation contextualisée des formes discursives. Une telle démarche aiderait les apprenants à mieux comprendre pourquoi et comment exprimer leur opinion de manière appropriée, tout en gérant la relation interpersonnelle dans des contextes informels.

## Bibliographie

- Aijmer, K. (1986). *Discourse variation and hedging* (p. 1-18). Brill.
- André, V. (2023). De la constitution d'un corpus d'interactions à son exploitation en didactique de l'oral. Quels outils et quel accompagnement pour des pratiques innovantes?. *Recherches en didactique des langues et des cultures, Les cahiers de l'Acedle*, (21-2), 1-18.
- Baude, O., Blanche-Benveniste, C., Calas, M.-F., Cappeau, P., Cordereix, P., Goury, L., Jacobson, M., De Lamberterie, I., Marchello-Nizia, C. et Mondada, L. (2006). *Corpus oraux, guide des bonnes pratiques 2006*. CNRS Editions, Presses Universitaires Orléans.
- Brown, P. & Levinson, S. C. (1987). *Politeness: some universals in language usage*. Cambridge University Press.
- Chen, R. (2024). *Exprimer son désaccord en français et en chinois dans des échanges informels : analyse contrastive d'interactions verbales* [Thèse de doctorat]. Université de Lille.
- Debaisieux, J.-M. (2005). Les corpus oraux : situation, exploitation linguistique, bilan et perspectives. *Solia [sciences cognitives, linguistique et intelligence artificielle / revue de linguistique]*, (19), 9.
- Debaisieux, J.-M. (2009). Des documents authentiques oraux aux corpus : un défi pour la didactique du fle. *Mélanges CRAPEL*, 31(1), 35-56.
- Debras, C. (2021). Multimodal profiles of je (ne) sais pas in spoken French. *Journal of Pragmatics*, 182, 42-62.
- Delahaie, J. (2009). Oui, voilà ou d'accord ? Enseigner les marqueurs d'accord en classe de FLE. *Synergies, Pays Scandinaves*, 4, 17-34.
- Delahaie, J. (2013). Constitution et exploitation de corpus d'interactions verbales pour le FLE : problèmes et programme. *Linx. Revue des linguistes de l'université Paris X Nanterre*, (68-69), 95-114.
- Delahaie, J., Canut, E., Balvet, A. & del Olmo, C. (2018). *Le projet DOC (Didactique, Oral, Corpus), constitution d'une base de données d'interactions orales pour l'enseignement du français langue étrangère*. Orléans, France.
- Delahaie, J., Rouanne, L. & Canut, E. (2022). Les verbes recteurs faibles dans l'expression de l'opinion en français et en espagnol. *Travaux de linguistique*, 8485(1), 131-157.
- Dostie, G. (2004). *Pragmaticalisation et marqueurs discursifs. Analyse sémantique et traitement lexicographique*. De Boeck Supérieur.
- Dostie, G. (2016). Le Corpus de français parlé au Québec (CFPQ) et la langue des conversations familiales : Exemple de mise à profit des données à partir d'un examen lexicosémantique de la séquence *je sais pas*. *Corpus*, (15).
- Gadet, F. & Guérin, E. (2008). Le couple oral / écrit dans une sociolinguistique à visée didactique. *Le français aujourd'hui*, 162(3), 21-27.
- Hallion, S. (2020). De la constitution d'un corpus de français oral à son exploitation pédagogique : enjeux, objectifs, limites et perspectives. In P. Dupont (éd.), *L'enseignement de l'oral en contexte francophone : pratiques et outils de formation* (pp. 51-66). Presse Universitaire du Midi.
- Kärkkäinen, E. (2010). Position and Scope of Epistemic Phrases in Planned and Unplanned American English. *Studies in Pragmatics*, 9, 203-236.
- Kerbrat-Orecchioni, C. (2001). *Les actes de langage dans le discours: théorie et fonctionnement*. Armand Colin.
- Kerbrat-Orecchioni, C. (2005). *Le discours en interaction*. Armand Colin.
- Kiparsky, P. & Kiparsky, C. (1970). Fact. Dans *Progress in Linguistics. A collection of Papers*.

The Hague.

Li, L. (李丽娟). (2015). 动词“看”“想”“说”“知道”为核心构成的话语标记研究[A Study on Discourse Markers Which Were Composed by The Verbs « look » « think » « say » and 'know'] [thèse de doctorat]. 华中师范大学[Central China Normal University].

Nguyen, A. & Tree, J. (2025). Pragmatic uses of I don't know, boosters, and hedges in text and talk. *Dialogue & Discourse*, 16, 68-90.

Nguyen, M. T. T. (2011). Learning to communicate in a globalized world: to what extent do school textbooks facilitate the development of intercultural pragmatic competence? *RELC Journal*, 42(1), 17-30.

Pekarek Doehler, S. (2016). More than an epistemic hedge: French je sais pas 'I don't know' as a resource for the sequential organization of turns and actions. *Journal of Pragmatics*, 106, 148-162.

Pichler, H. (2007). Form-function relations in discourse: The case of I DON'T KNOW. *Newcastle Working Papers in Linguistics*, 13, 174-187.

Rémi-Giraud, S. (1986). Étude comparée du fonctionnement sémantique et syntaxique des verbes savoir et connaître. In S. Rémy-Giraud & M. Le Guern (éds.), *Sur le verbe* (pp. 169-306). Presses Universitaires de Lyon.

Scheibman, J. (2000). I dunno: a usage-based account of the phonological reduction of don't in American English conversation. *Journal of Pragmatics*, 32(1), 105-124.

Spencer-Oatey, H. & Jiang, W. (2003). Explaining cross-cultural pragmatic findings: Moving from politeness maxims to sociopragmatic interactional principles (SIPs). *Journal of Pragmatics*, 35, 1633-1650.

Tao, H. (陶红印). (2003). 从语音、语法和话语特征看“知道”格式在谈话中的演化 [Évolution du verbe « savoir » dans la conversation à partir de caractéristiques phonologiques, grammaticales et discursives]. *中国语文 [Langue Chinoise]*, (04), 291-302+383.

Ten Berge, T. & Van Hezewijk, R. (1999). Procedural and Declarative Knowledge: An Evolutionary Perspective. *Theory & Psychology*, 9(5), 605-624.

Terkourafi, M. (2002). Politeness and formulaicity: Evidence from Cypriot Greek. *Journal of Greek Linguistics*, 3(1), 179-201.

Traverso, V. (1996). *La conversation familière: analyse pragmatique des interactions*. Presses Universitaires de Lyon.

Tsui, A. B. M. (1991). The pragmatic functions of I don't know. *Text - Interdisciplinary Journal for the Study of Discourse*, 11(4), 607-622.

Tutin, A. (2019). Phrases préfabriquées des interactions : quelques observations sur le corpus CLAPI. *Cahiers de Lexicologie*, (114), 63-91.

Weatherall, A. (2011). I don't know as a Prepositioned Epistemic Hedge. *Research on Language and Social Interaction*, 44(4), 317-337.

Xu, R. & Zhang, C. (2016). Enjeux des corpus oraux pour l'enseignement universitaire du français langue étrangère en Chine – L'exemple de la base de données CLAPI. *Synergies Chine*, (11), 177-188.

Yin, S. (殷树林) & Yin, L. (殷璐璐). (2017). 从真实会话语料看“知道”的用法——兼与陶红印先生商榷[the usage of « zhidao » in real conversations: a discussion with mr. tao hongyin]. *湖南科技大学学报(社会科学版)[Journal de l'Université des sciences et technologies de Hunan (version sciences sociales)]*.