

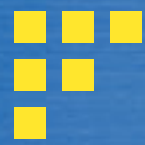
Aide au dialogue en langue étrangère



LIG
GETALP

Achille Falaise
achille.falaise@imag.fr
Doctorant CIFRE





Plan

- Introduction / contexte
- Etat de l'art
- Caractéristiques du dialogue en langue étrangère
- Problématique
- Vers une solution
- Mise en œuvre
- Conclusion



Contexte

La « mondialisation »

- **Accroissement des situations de multilinguisme**
 - Multinationales, projets internationaux (recherche, informatique)
 - « Corporations » transnationales
 - aéronautique, marine
 - Voyages
 - Tourisme international 1950: 25M → 2000: 500M
 - Expatriation (4% de la population mondiale)
 - Internet
 - Cloisonnement linguistique



Contexte

Les pidgins

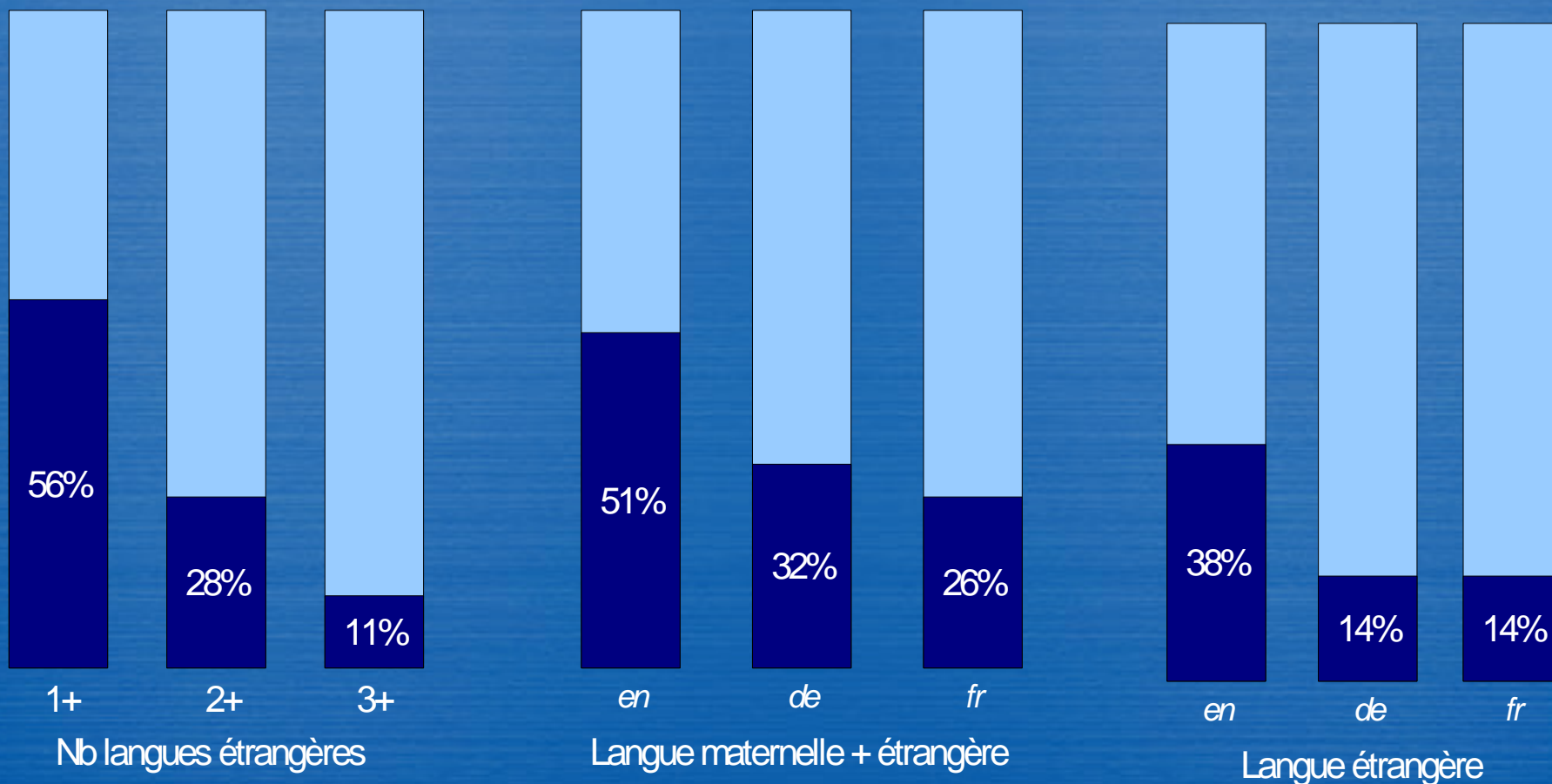
- Langue véhiculaire/utilitaire
 - Diplomates, scientifiques, lettrés, marchands, etc...
 - Quelques exemples en Europe:
 - Antiquité: grec *koinè*
 - Moyen-Âge: latin médiéval
 - XI-XIXème siècle: *lingua franca*
 - Fin XXème siècle: *global english, airspeak, seaspeak*

Contexte

Etude de la CE: les langues étrangères dans l'UE en 2006

+

+





Etat de l'art

Les outils de dialogue parlé multilingue

- Livres de phrases
 - Phraselator (VoxTec): unidirectionnel
 - S-Minds (Sedha): bidirectionnel
 - Interpreter (SpeechGear): phrases à trous, bidirectionnel
 - PSP Talkman (Sony): unidirectionnel, interactif, ludique
- Traduction automatique
 - Interact (SpeechGear)
 - Converser (Spoken): interactif

Etat de l'art

Les outils de dialogue parlé multilingue

- Livres de phrases
 - **Phraselator (VoxTec)**: unidirectionnel
 - S-Minds (Sedha): bidirectionnel
 - Interpreter (SpeechGear)
 - PSP Talkman (Sony): unidirectionnel
- Traduction automatique
 - Interact (SpeechGear)
 - Converser (Spoken): interactif



Etat de l'art

Les outils de dialogue parlé multilingue

■ Livres de phrases

- Phraselator (VoxTec): unidirectionnel
- S-Minds (Sedha): bidirectionnel
- Interpreter (SpeechGear)
- PSP Talkman (Sony): unidirectionnel

■ Traduction automatique

- **Interact** (SpeechGear)
- Converser (Spoken): interactif



el



Etat de l'art

Les outils de dialogue parlé multilingue

- Livres de phrases
 - Phraselator (VoxTec): unidirectionnel
 - S-Minds (Sedha): bidirectionnel
 - Interpreter (SpeechGear): phrases à trous, bidirectionnel
 - PSP Talkman (Sony): unidirectionnel, interactif, ludique
- Traduction automatique
 - Interact (SpeechGear)
 - Converser (Spoken): interactif



Etat de l'art

Les outils de dialogue écrit multilingue

- Livres de phrases
 - Phraselator (VoxTec): unidirectionnel
 - S-Minds (Sedha): bidirectionnel
 - Interpreter (SpeechGear): phrases à trous, bidirectionnel
 - PSP Talkman (Sony): unidirectionnel, interactif, ludique
- Traduction automatique
 - Interact (SpeechGear)
 - Converser (Spoken): interactif



Etat de l'art

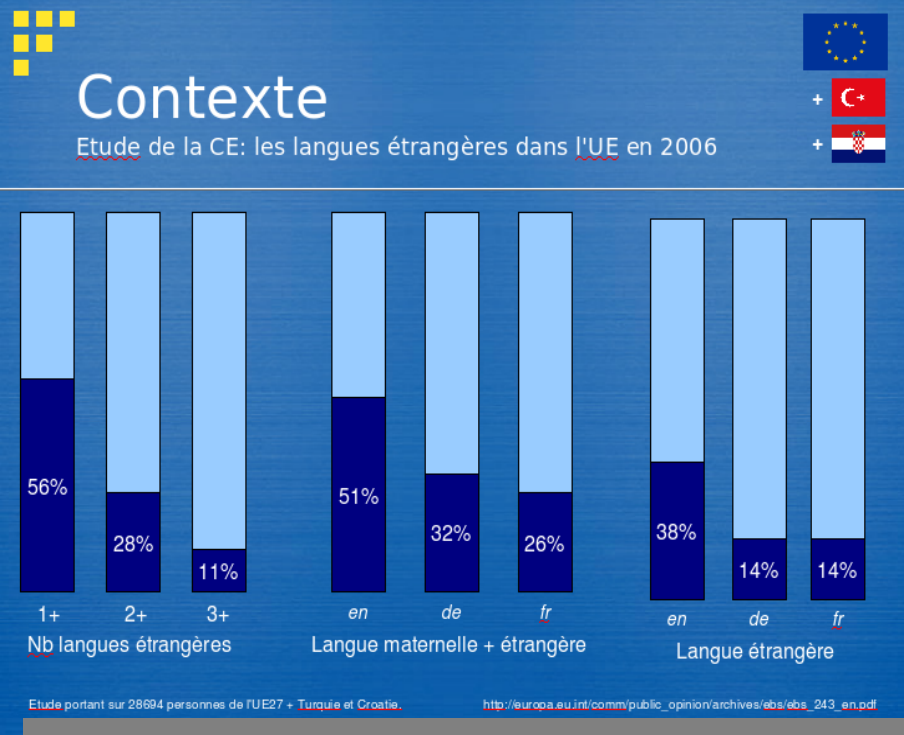
Critique 1/2

- Cas d'école: les locuteurs qui parlent des langues différentes

Etat de l'art

Critique 1/2

- Cas d'école: les locuteurs qui parlent des langues différentes



- Mais... pas systématique
- Pas conçu pour les interlocuteurs partageant une langue pivot



Etat de l'art

Critique 2/2

- Pas d'interaction prévue avec l'interlocuteur
 - Pourtant il pourrait aider:
 - s'il est plus doué: proposer du vocabulaire, des éléments de syntaxe, etc...
 - dans tous les cas: indiquer ce qu'il a compris et ce qui pose problème
 - Médiation de la machine → filtrage
 - besoin de retours, d'empathie (fonction phatique)
 - besoin d'authenticité
(cf. tchat: émoticons + onomatopées = 7%)
 - Pas ou peu de contrôle sur ce que la machine produit



Caractéristiques du dialogue en L2

- Didactique des langues
 - Ressources (corpus)
- Anglais langue internationale (*English as lingua franca*)
 - Anna Mauranen (Tampere), Christiane Meierkord (Erfurt), Barbara Seidlhofer (Vienne)
- Caractéristiques et problèmes principaux :
 - dialogues superficiels
 - malentendus
 - co-construction du sens (*wadan*)

Caractéristiques du dialogue

- Didactique des langues
 - Ressources (corpus)
- Anglais langue internationale
 - Anna Mauranen (Talbot)
- Caractéristiques et problèmes du dialogue
 - dialogues superficiels
 - malentendus
 - co-construction du sens (*negotiation*)

*NIB: pourquoi est ce que tu as choisi l' université de Salford?
*C01: um j' aime Manchester.
*C01: et je trust@s euh j' ai voulu étudier à Manchester et à Salford.
*C01: j' aime le cours.
*NIB: tu viens de quelle partie de l' Angleterre?
*C01: um pas d' Angleterre.
*C01: du pays de Galles.
*NIB: du pays de Galles?
*C01: oui.
*NIB: euh d' ou au pays de Galles?
*C01: euh le sud.
*NIB: le sud?
*C01: oui de Swansea.
*NIB: um est ce que tu as le mal du pays ici?
*C01: um xxx oui beaucoup.
*NIB: tu rentres pour Noël?
*C01: euh oui.
*NIB: ok ou est ce que tu habites à Salford?
*C01: euh j' habite à halls.
*NIB: halls?
*C01: halls.

Salford corpus. Investigators: R D Hawkins and R J Towell. Data was collected by N P Bazergu.



Caractéristiques du dialogue en L2

- Didactique des langues
 - Ressources (corpus)
- Anglais langue internationale (*English as lingua franca*)
 - Anna Mauranen (Tampere), Christiane Meierkord (Erfurt), Barbara Seidlhofer (Vienne)
- Caractéristiques et problèmes principaux :
 - dialogues superficiels
 - malentendus
 - co-construction du sens (*wadan*)

NS: what are you mumbling?

NNS: that song (0.6) hhhhm. hh. what's that woman's name? that singer (0.2) **the chocolate woman !**

NS: areta franklin

NNS: no::::! she's not black!

NS: you said she was black!

NNS: no I did not!

NS: yes you did!

NNS: NO I DIDN'T

NS: yes you did. you said that chocolate woman.

NNS: yes, the one that sings that your love is better than chocolate

NS: (0.3) sarah madofn? she's white

NNS: I never said she was black!

NS: yes (.) you (.) did

NNS: NO I didn't ((frustrated)) (3.4)

NS: ((laughs))

NNS: whats so funny↑ ((quite angry))

NS: honey, a chocolate woman IS a black woman.

NNS: oh (0.4) how am I supposed to know that?

Marta Gonzalez-Lloret, « Reconstructing NS/NNS communication », in *The Consequences of Mobility*, 2005

<http://ruc.dk/isok/skriftserier/mobility/mobility2/Gonzalez-Lloret/>



Caractéristiques du dialogue en L2

- Didactique des langues
 - Ressources (corpus)
- Anglais langue internationale (*English as lingua franca*)
 - Anna Mauranen (Tampere), Christiane Meierkord (Erfurt), Barbara Seidlhofer (Vienne)
- Caractéristiques et problèmes principaux :
 - dialogues superficiels
 - malentendus
 - co-construction du sens (*wadan*)



Caractéristiques du dialogue en L2

- Didactique des langues
 - Ressources (corpus)
- Anglais langue internationale (*English as lingua franca*)
 - Anna Mauranen (Tampere), Christiane Meierkord (Erfurt), Barbara Seidlhofer (Vienne)
- Caractéristiques et problèmes principaux :
 - dialogues superficiels
 - malentendus
 - co-construction du sens (*wac*)

Christiane Meierkord, « Interpreting successful lingua franca interaction. An analysis of non-native-/non-native small talk conversations in English. », *Linguistik online* 5, 1/00, 2000.
http://www.linguistik-online.de/1_00/MEIERKOR.HTM



Problématique

- Quel outil spécifique pour la conversation en langue étrangère ?
 - Tenir compte des spécificités de cette modalité.
 - Impliquer l'interlocuteur.
 - Réduire le « filtrage ».
 - Limiter les contraintes pour l'utilisateur.
 - Accroître la confiance dans l'énoncé produit.
 - Proposer sans imposer.



Vers une solution

Scénario

- Prosodie
 - Grands comptes
 - Filiales aux Etats-Unis et en Espagne, prospect au Japon

- Discussions d'affaires
 - Testable en interne et commercialisable
 - Caractéristiques
 - Anglais
 - Rendez-vous, commandes, etc...
 - Deux situations :
 - Locuteur natif + locuteur non-natif
 - Locuteur non-natif + locuteur non-natif



Vers une solution

Tentative 1/2

- Sous-titrage « temps réel »
 - Problèmes avec la reconnaissance vocale
 - Nécessité de dicter
 - Problème des accents
 - Vocabulaire réduit
 - Problèmes d'ergonomie
 - Gestion du différé
 - Interaction avec les sous-titres
 - Risque d'aboutir à quelque chose d'inutilisable



Vers une solution

Tentative 1/2

- Sous-titrage « temps réel »
 - Problèmes avec la reconnaissance vocale
 - Nécessité de dicter
 - Problème des accents
 - Vocabulaire réduit
 - Problèmes d'ergonomie
 - Gestion du différé
 - **Interaction avec les sous-titres**
 - Risque d'aboutir à quelque chose d'inutilisable



Vers une solution

Tentative 2/2

- Aide à la conversation *écrite* (→ tchat)
 - Instrumentation
 - mesure de la correction
 - correction orthographique
 - nombre d'occurrences sur le web
 - concordancier-web pour vérifier l'usage
 - mesure de la difficulté
 - propositions de vocabulaire en fonction du contexte



Vers une solution

Tentative 2/2

- Outils
 - lexicaux: livre d'images, dictionnaire multilingue
 - traduction automatique
- Retouche / interaction
 - affichage et résolution interactive des ambiguïtés (cf. Supeeti)
 - interventions sur le texte
 - lancement de sous-dialogues (*wadan*)

→ *surtout des problèmes d'ergonomie*



Mise en œuvre

Contraintes

- Contraintes sur l'interface
 - Du point de vue utilisateur
 - multilingue
 - facile à utiliser, conforme aux standards actuels d'ergonomie
 - interface attrayante
 - facile à installer, multi-plateforme
 - Du point de vue développeur
 - approche haut-niveau
 - aisément modifiable
 - nombreux composants → approche modulaire



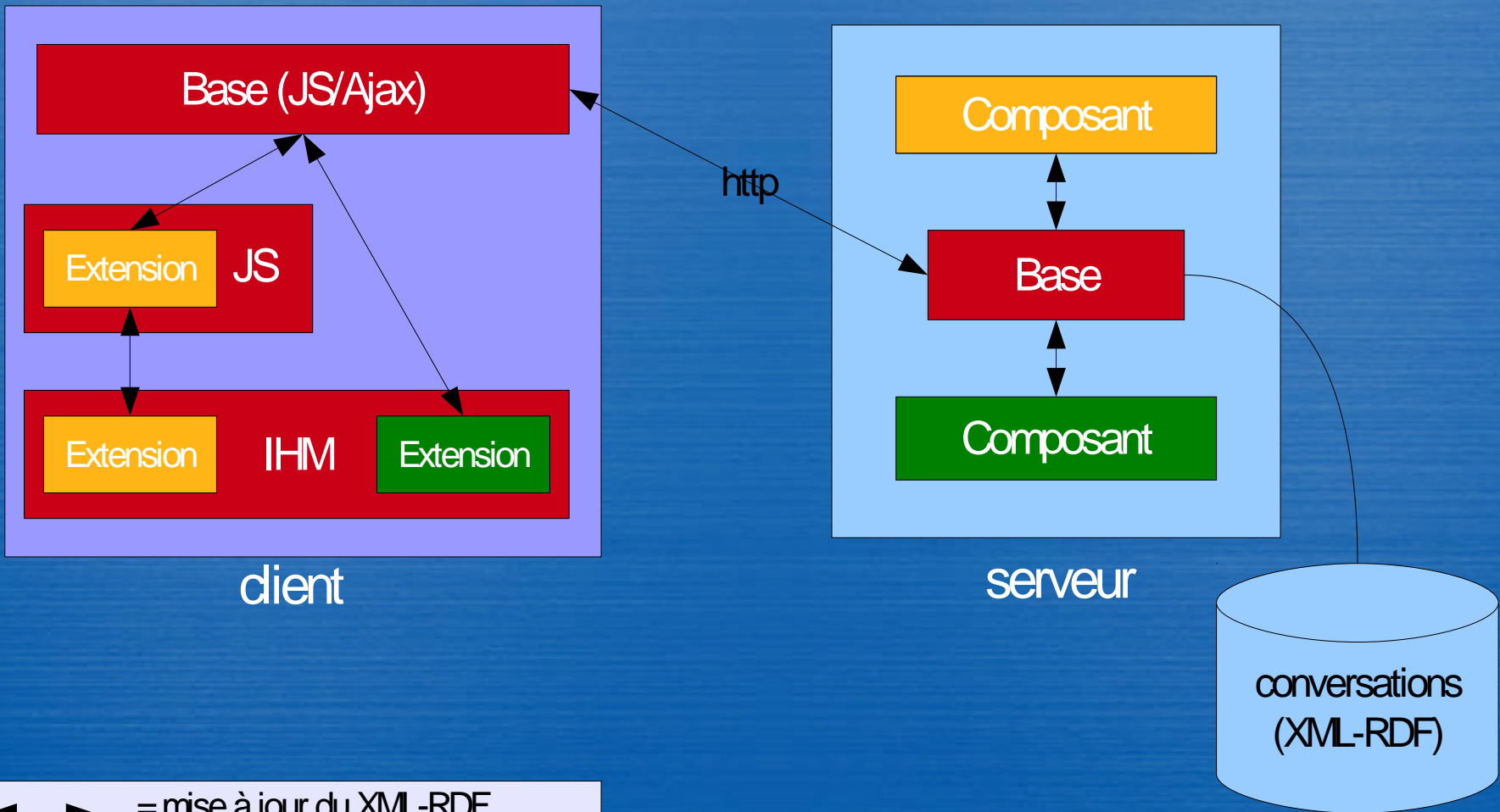
Mise en œuvre

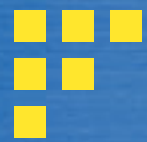
Choix techniques

- Application web
 - Séparation nette interface / données / traitements
 - Communication par tableau blanc (cf. Verbmobil)
 - Client: basé sur le framework Mozilla
 - XPFE (*Cross Platform Front-End*): XUL, XHTML, CSS, JavaScript
 - *templates*, extensions, composants riches et extensibles
 - librairie XPCOM: gestion native du format de données RDF-XML
 - à la fois site web - standalone - extension Mozilla
 - Données: RDF-XML
 - Serveur:
 - un servlet java par composant

Mise en œuvre

Architecture, exemple avec 2 composants





Conclusion

- Point de vue de « linguiste » sur l'aide à la conversation
 - Appréhender les problèmes réels (corpus)
- Recherche de solutions
 - ... du point de vue de l'ergonomie
 - ... jouant sur l'interaction et la motivation des locuteurs pour faire avancer les choses
 - ... réalisables



Conclusion

■ Perspectives

- Prototype: terminer gestion des extensions (+documentation)
- Formaliser les hypothèses
 - concernant le dialogue écrit en langue étrangère et ses problèmes
 - concernant les moyens informatiques de résoudre ces problèmes
- Prototype: ajouter les extensions requises
 - dont un *plug-in* d'entrée par reconnaissance vocale
- IHM: intégrer les extensions entre-elles et avec la base
- Expérimentation, collecte de données
- Analyse des données, critique et réactualisation des hypothèses
- Spécification d'une seconde version