

# Analyse multimodale de la cohésion de groupe

Reshmashree B Kantharaju, Caroline Langlet, Mukesh Barange, Chloé Clavel, Catherine I Pelachaud

▶ **To cite this version:**

Reshmashree B Kantharaju, Caroline Langlet, Mukesh Barange, Chloé Clavel, Catherine I Pelachaud. Analyse multimodale de la cohésion de groupe. Workshop sur les Affects, Compagnons artificiels et Interactions, CNRS, Université Toulouse Jean Jaurès, Université de Bordeaux, Jun 2020, Saint Pierre d'Oléron, France. hal-02933466

**HAL Id: hal-02933466**
**https://inria.hal.science/hal-02933466**

Submitted on 8 Sep 2020

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

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

# Analyse multimodale de la cohésion de groupe*

Reshmashree B Kantharaju
ISIR, Sorbonne Université
bangalore_kantharaju@isir.upmc.fr

Caroline Langlet
ISIR, Sorbonne Université
langlet@isir.upmc.fr

Mukesh Barange
ISIR, Sorbonne Université
barange@isir.upmc.fr

Chloé Clavel
LTCI, Télécom Paris-Tech
chloe.clavel@telecom-paris.fr

Catherine Pelachaud
CNRS-ISIR, Sorbonne Université
pelachaud@upmc.fr

## ABSTRACT

La cohésion de groupe est un concept permettant de décrire la tendance que les membres d'un groupe ont de s'engager dans les différentes tâches de ce groupe, ainsi que l'attrait interpersonnel existant entre eux. Cet article décrit une analyse multimodale de la cohésion de groupe dans un corpus d'interactions multipartites. 16 segments de deux minutes chacun ont été utilisés pour mener cette analyse. Trois niveaux de modalité sont pris en compte : des indices sociaux non-verbaux, les actes de dialogue et les interruptions. Dans un premier temps, l'analyse se concentre d'abord sur chacun de ces niveaux pris individuellement, pour, dans un second temps, les observer conjointement. Les résultats obtenus montrent que les co-occurrences de rires et d'interruptions sont plus importantes dans des segments ayant un score de cohésion élevé. Nous observons également que les actes de dialogue et les hochements de tête n'ont pas d'impact sur le niveau de cohésion lorsqu'ils sont observés individuellement. Néanmoins, une incidence sur le niveau de cohésion est observée lorsqu'ils sont analysés conjointement. Globalement, notre étude montre que la prise en compte d'indices multimodaux est cruciale pour une analyse précise de la cohésion de groupe.

## KEYWORDS

Cohésion, Actes de Dialogue, Comportement Non-verbaux, Interruptions

## 1 INTRODUCTION

Les conversations de groupe sont une forme essentielle de la communication humaine. Les être humains peuvent être amenés à discuter, à prendre des décisions ou à échanger des idées au sein de groupes dont les formes varient (par exemple des réunions, des conférences, des conseils, etc). La littérature en sociologie et en psychologie livre un ensemble d'études sur différents aspects des dynamiques de groupe [24]. Si de nombreuses questions de recherche concernent les comportements humains dans les groupes, notre recherche s'intéresse plus précisément à la cohésion au sein des groupes. La cohésion concerne la tendance que les membres d'un groupe ont à partage un lien ou un attrait qui les amènent à travailler ensemble [17]. Un groupe montrant une forte cohésion peut être défini comme un groupe dont les membres cherchent à rester ensemble et manifestent un sentiment de solidarité et d'harmonie [35]. Ce phénomène émerge au cours du temps, dans les équipes de travail notamment [42]. Plusieurs travaux montrent une correspondance entre la cohésion d'un groupe, ses performances et sa satisfaction [4].

Une estimation automatique du niveau de cohésion de groupe peut s'avérer utile pour analyser automatiquement, par exemple, des corpus de réunions. Cet article est une première étape vers le développement d'un modèle computationnel d'estimation de la cohésion dans dans des interactions multipartites humain-humain et humain-agent. Pour développer un modèle de ce type, il est nécessaire de prendre en compte des facteurs correspondant non seulement à des informations de haut niveau, par exemple la gestion des tours de parole et les actes de dialogue, mais également à des informations de bas niveau, tels que les comportements non-verbaux. Dans un objectif de clarté, nous référons aux comportements correspondant à des informations de bas niveau – regard, hochement de tête et rire – en utilisant l'expression *indices sociaux non-verbaux*. Cet article présente ainsi une analyse préliminaire de la manière dont comportements multimodaux de haut et de bas niveaux sont associés à la cohésion de groupe dans un corpus d'interaction humain-humain. Notre but est ainsi de mettre en évidence les principales caractéristiques de la cohésion de groupe.

Dans les interactions multipartites, i.e les interactions avec plusieurs participants, les participants communiquent et se coordonnent grâce à un certain nombre de comportements verbaux et non-verbaux. Ils prennent la parole et généralement ces prises de parole commencent et finissent de manière fluide avec un court temps de pause entre eux. Néanmoins, des chevauchements, des interruptions et des silences peuvent apparaître [43]. La littérature sur l'estimation de la cohésion a ainsi montré une forte corrélation entre le niveau de cohésion et le nombre d'interruptions. Pour cette raison, nous avons souhaité intégrer les interruptions à notre analyse. Au total, trois niveaux de modalité ont été considérés : les indices sociaux non-verbaux, les actes de dialogue et les interruptions. Chaque niveau est étudié individuellement afin comprendre son impact sur le niveau de cohésion perçu. Ces niveaux sont ensuite analysés conjointement pour analyser la manière dont leur combinaison impacte le niveau de cohésion.

Dans le section 2, nous présentons la cohésion de groupe telle qu'elle est décrite en psychologie, ainsi que les comportements de communication qui peuvent lui être associés sur un plan dialogique. Dans la section 2.3, nous décrivons notre approche en trois niveaux – indices sociaux non-verbaux, actes de dialogue et interruptions. Ensuite, dans la section 3, nous présentons les données et les annotations que nous avons utilisées pour notre analyse. La section 4 présente et discute les résultats de l'analyse de chaque niveau pris

*Cet article est une traduction légèrement modifiée de l'article R. B Kantharaju et. al., *Multimodal Analysis of Cohesion in Multi-party Interactions* soumis à la conférence LREC 2020.

individuellement. Enfin, la section 5 propose une analyse conjointe de ces comportements en relation avec le niveau de cohésion.

## 2 COHÉSION

Plusieurs définitions de la cohésion ont été proposées dans des contextes aussi spécifiques que les équipes de sport [15] ou les thérapies de groupe [7]. L'une des premières définitions a été proposée par Festinger et. al. Ils définissent la cohésion comme l'ensemble de toutes les forces qui agissent sur les membres pour rester au sein du groupe [23]. D'autres chercheurs ont proposé des définitions incluant l'attrait [1] ou l'engagement soit envers le groupe soit envers les tâches prises en charge par le groupe [38],[27]. Néanmoins, ces définitions considèrent la cohésion comme un phénomène unidimensionnel.

Carron et. al. définissent la cohésion comme un processus multidimensionnel et dynamique se caractérisant par la tendance que les membres du groupe ont à faire corps et à rester unis pour accomplir leurs objectifs [14]. Sur la base de cette définition, les auteurs proposent un modèle multi-dimensionnel qui distingue d'une part la dimension du groupe et celle de l'individu, et, d'autre part, la dimension sociale et la dimension de tâche [16]. La distinction groupe-individu indique que la cohésion résulte à la fois du désir de la part de chacun de rester impliqué dans le groupe et de faire unité (intégration dans le groupe, IG), et de l'attrait personnel de chacun à être un membre du groupe (attrait interpersonnel vers le groupe, AIG). La distinction tâche-sociale permet de différencier la cohésion concernant les tâches à effectuer par le groupe et de celles reposant sur des aspects sociaux. La cohésion sociale peut se mesurer grâce à l'attrait interpersonnel existant entre les membres du groupe, tandis que la cohésion de tâche se mesurera grâce au degré d'implication de chaque membre dans les tâches communes à accomplir. Au total, quatre dimensions sont considérées : AIG-tâche, AIG-sociale, IG-tâche and IG-sociale. Braaten propose un modèle à 5 facteurs pour analyser la cohésion dans des thérapies de groupe : attrait et capacité à créer des liens, soutien et attention, écoute et empathie, révélation de soi et retour d'expérience, performance et accomplissement d'objectif [7]. Carless and De Paola [12] proposent également un modèle à trois dimensions : cohésion de tâche, cohésion sociale et attrait pour le groupe.

### 2.1 La Cohésion dans une Perspective Dialogique

Pour analyser la cohésion dans une perspective dialogique, il est nécessaire d'observer les comportements relevant l'attrait interpersonnels des locuteurs pour le groupe. En linguistique, le concept de cohésion ne réfère pas à la cohésion de groupe telle qu'elle est définie en psychologie, mais à la cohésion du discours lui-même. Dans [44], l'auteur décrit la cohésion comme le principe selon lequel l'interprétation de certains éléments du discours dépend de l'interprétation d'autres éléments. Un discours montre de la cohésion lorsqu'il fonctionne comme une unité. La cohésion verbale peut se faire grâce à des relations entre les différents éléments du discours, comme par exemple les relations de co-référence et de similarité lexicale. Ainsi, elle concerne la manière dont les différentes parties du discours s'articulent et créent un tout cohérent et signifiant sur le plan sémantique et pragmatique. Pour comprendre comment les locuteurs interagissent en cohésion et comment la cohésion de groupe se reflète dans leur comportement verbal, il est nécessaire de se tourner vers les études d'analyse du dialogue. Ces dernières décrivent le dialogue comme une activité commune, comme une tâche effectuée en collaboration [34]. La cohésion n'est pas explicitement nommée dans ce type de travaux, néanmoins nous formulons l'hypothèse que certains comportements verbaux liés au phénomène de la collaboration dialogique peuvent être révélateurs du niveau de cohésion de groupe. Ce sont ces comportements que nous introduisons ci-dessous :

*Alignement.* Les études sur l'alignement s'intéressent à la manière dont les locuteurs s'adaptent au niveau verbal pour diminuer ou accentuer les différences sur le plan communicatif et social. L'alignement regroupe un certain nombre de comportements verbaux et non-verbaux. Concernant l'alignement verbal, la plupart des travaux étudient le dialogue comme un processus d'imitation et de coordination et analysent la manière dont l'alignement peut provoquer une connexion sociale entre les locuteurs. Plusieurs études ont montré que les participants d'un dialogue s'alignent spontanément et ce à différents niveaux : au niveau lexical, au niveau syntaxique et au niveau sémantique. Reitter et. al. [40] ont montré que les locuteurs réemploient autant les unités lexicales que les structures syntaxiques utilisées dans les tours de parole précédents. Caractéristique naturelle des dialogues humain-humain [37], l'alignement verbal peut être utilisé dans les interactions humain-agent pour améliorer les compétences de communication de l'agent [11]. Puisque l'alignement est en rapport avec la coordination et la connexion sociale, notre hypothèse est qu'il peut être un indicateur, dans le contenu verbal, de la cohésion existant entre les participants du dialogue.

*Synergie interpersonnelle.* Si l'alignement concerne la manière dont les participants se coordonnent en utilisant des répétitions tour de parole après tour de parole, la synergie interpersonnelle concerne davantage la manière dont les participants vont compléter mutuellement leurs énoncés pour créer une conversation cohérente et faisant sens [34]. Le concept de synergie interpersonnelle permet d'appréhender la manière dont l'interdépendance entre locuteurs repose sur une forme de complémentarité [25]. Fusaroli et. al. considèrent que, dans une conversation, les routines pré-existantes et négociées localement créent une interdépendance entre locuteurs au niveau de leurs comportements verbaux [25]. Les routines sont des modèles de comportements et reposent sur une dynamique de complémentarité. La complémentarité dans le dialogue apparaît au niveau des séquences de tour de parole, comme par exemple les paires adjacentes. Ainsi, aux questions doivent normalement succéder des réponses, et non d'autres questions. De même les invitations sont généralement suivies par une acceptation ou un rejet [25]. Dans cet article, notre but est d'étudier les phénomènes verbaux qui sont en lien avec la synergie interpersonnelle. Puisque les séquences de tours de parole reposent sur la complémentarité entre les participants du dialogue, notre hypothèse est que celles-ci peuvent être un indice du niveau de cohésion dans le groupe.

*Turn Taking et Interruptions.* Le succès d'une interaction multipartites repose sur la coordination des membre du groupe pour

prendre la parole [6]. Dans les interactions dialogiques, le concept de *turn taking* désigne cette capacité qu'ont les participants à faire alterner la prise de parole, pour laisser s'exprimer tout participant qui en montre l'intention. Néanmoins, dans les interactions multipartites, des chevauchements peuvent apparaître lorsque plusieurs participants essayent de s'exprimer simultanément [29]. Ces chevauchements en tours de parole peuvent être considérés comme la caractéristique tant d'une coopération [45] que d'un conflit [49] au sein du groupe. La violation des règles basiques de *turn taking* peut générer une interruption, l'un des participant coupant la parole à un autre avec un nouvel énoncé. Selon son contenu, l'interruption peut se révéler coopérative ou disruptive [33]. Une interruption coopérative exprimera le soutien, l'accord, une demande de clarification ou permettra de clore l'énoncé du locuteur courant. Une interruption disruptive exprimera le rejet, un changement thématique ou un désaccord. Les interruptions et les changements de tour de parole ont été étudiés pour analyser le comportement des participants d'un groupe. Par exemple, G. W. Beattie étudie les interruptions selon le genre et le statut des participants [5]. Les interruptions ont également été utilisées pour étudier la relation entre genre et dominance [45]. Les résultats montrent que les interruptions ne sont pas nécessairement une marque de dominance dans les interactions de groupe. Elles se révèlent également être plus fréquentes dans les conversations multipartites que dans les conversations dyadiques [5]. Dans les interactions multipartites, les participants ont tendance à se relayer, à parler plus souvent car le locuteur courant peut céder la parole à plusieurs auditeurs. De plus, dans ce type d'interaction, il n'est pas évident que seulement deux personnes (celle qui interrompt et celle qui est interrompue) participent à une interruption. Par exemple, d'autres participants peuvent interrompre la personne ayant la parole et commencer à parler à quelqu'un d'autre [3, 39]. Cafaro *et* al. ont observé les effets des interruptions dans les conversations dyadiques et ont montré que le type d'interruption – coopérative ou disruptive – a un impact sur le perception des attitudes interpersonnelles [9].

## 2.2 Analyse automatique de la cohésion

Plusieurs études ont présentés des techniques pour collecter et analyser les données relatives à la cohésion. Par exemple, des badges sociométriques ont été utilisés pour estimer la cohésion en se basant sur la proximité temporelle, la durée de l'interaction et la fréquence [50]. Hung et. al., [30] ont travaillé sur l'estimation et l'annotation du niveau de cohésion. Pour cela, les auteurs font appel à des observateurs extérieurs. Les résultats révèlent que la caractéristique la plus pertinente est le nombre total de pauses entre chaque tour de parole. Ils montrent également une forte corrélation entre le niveau de cohésion et le *turn taking*. Leur étude souligne également que l'extraction automatique d'indices comportementaux s'avère utile pour estimer automatiquement le niveau de cohésion. S. Fang et C. Achard étudient la relation entre la cohésion et la personnalité des participants [22]. Les résultats montrent une forte corrélation entre l'amabilité (en tant que trait de personnalité) et la cohésion. De même, la cohésion se révèle être liée au *turn taking* et à la variation du *speech energy*. Wang et. al. proposent trois catégories pour appréhender un groupe : cohésion, division et mixité de l'interaction [48]. Pour l'analyse de la cohésion, ils se réfèrent également à des phénomènes tels que les marqueurs du discours ou les disfluences. Ils montrent que les interactions avec cohésion impliquent des expressions d'accord et un alignement avec des désaccords mineurs et d'autres formes de refus. L'étude de Nanninga et. al. proposent un calcul du niveau de cohésion reposant sur une analyse de contenu, i. e. une analyse de la *mimicrie* linguistique et para-linguistique et de la convergence [36]. Les résultats révèlent l'utilité de l'exploitation de la mimicrie paralinguistique pour l'estimation de la cohésion sociale. Celle-ci en effet apparaît s'exprimer davantage par des comportements acoustiques non-verbaux que le cohésion de tâche.

## 2.3 Notre approche

Dans cet article, nous analysons le lien entre les comportements verbaux et non-verbaux, et la cohésion de groupe dans des interactions multipartites. Pour cela, nous adaptons une approche multi-niveaux où chaque niveau correspond à un type de comportement. Tout d'abord, nous étudions chacun de ces niveaux individuellement pour estimer leur impact propre sur la perception de la cohésion. Ensuite, nous proposons une analyse par combinaison des différents niveaux pour définir l'impact de celle-ci sur la perception de la cohésion. Comme mentionné précédemment, les trois niveaux considérés sont les suivants : les indices sociaux non-verbaux, les actes de dialogues et les interruptions. Puisque notre but est de proposer un modèle computationnel pour l'estimation de la cohésion, notre analyse se concentre sur des comportements détectables semi-automatiquement qui sont annotés dans un corpus d'interactions multipartites.

*Indices sociaux non-verbaux.* Les comportements non-verbaux tels que le regard, les expressions faciales, les gestes et les positions du corps, etc. sont des indicateurs de l'attitude d'un individu dans toute situation sociale donnée [41]. Ils donnent des informations concernant les affects, l'état mental, la personnalité de ce dernier [47]. Tandis que de nombreux travaux proposent des analyses détaillées de caractéristiques telles que la prosodie ou l'énergie visuelle pour mesurer la cohésion, ils n'analysent pas les signaux sociaux en eux-mêmes. Par conséquent, pour cette analyse préliminaire, nous nous intéressons au *gaze*, aux *head nods*, aux *facial action units* et au *rire*. Puisque la cohésion est en relation avec les liens sociaux, les *feedback* et le *support*, notre hypothèse est que les comportements exprimant ces caractéristiques, comme le *mutual gaze*, les *head nods*, les rires ou les sourires sont fréquents dans des segments ayant un score élevé de cohésion. Dans notre analyse, nous nous intéressons également à la présence de l'*action unit* AU4 i. e., **brow lowerer** qui est souvent assimilé à l'expression d'une émotion négative e. g., colère, dégoût [20].

*Les actes de dialogue.* Comme expliqué dans la section 2), d'un point de vue dialogique, deux types de processus interpersonnels peuvent être en lien avec la cohésion de groupe i. e., *l'alignement* et *la synergie interpersonnelle*. Ces processus impliquent des comportements très différents : vocabulaire partagé, répétitions lexicales et syntaxiques pour l'alignement et routines et paires adjacentes pour la synergie interpersonnelle. Pour une première étape dans l'analyse des comportements verbaux en lien avec la cohésion, nous limitons

notre étude à celle de la synergie interpersonnelle. Les actes de dialogue sont des éléments essentiels à la construction des séquences structurées de tours de parole impliquées par la synergie interpersonnelle. Notre analyse repose sur l'étude d'un corpus d'interaction où les actes de dialogue sont annotés. Le schéma d'annotation sera présenté dans la section 3. Nous avons décidé de nous appuyer sur ces annotations inspirées par le schéma d'annotation de DIT++ [8]. Il apparaît en effet pertinent de s'appuyer sur une classification largement diffusée et connue des actes de dialogue tel que DIT++.

*Turn Taking et interruptions.* *Turn Taking* et interruptions sont des éléments fondamentaux pour un fonctionnement efficace d'une interaction de groupe. Les interruptions ne sont pas nécessairement dyadiques par nature dans une interaction en groupe. Les travaux présentés dans la section 2 montrent les effets du *turn taking* et des interruptions dans les interactions de groupe et mettent en lumière la manière dont les participants se comportent pendant ces interactions. Néanmoins, il n'existe que quelques études se consacrant à l'analyse de la cohésion. Par conséquence, cet article a pour objectif d'étudier la relation que le *turn taking* et les interruptions entretiennent avec la cohésion dans les interactions multipartites. Nous formulons l'hypothèse que le nombre de tours de parole, de chevauchements entre les tours de parole et d'interruptions est plus important dans les segments dont le score de cohésion est élevé.

## 3 DATASET

Cette section présente le corpus et les annotations que nous avons utilisés pour notre analyse. Le corpus *Augmented Multiparty Interaction* (corpus AMI) [13] regroupe 100 heures d'enregistrements multimodaux où quatre participants interagissent en suivant un scénario de réunion de travail. Le corpus a été annoté au niveau de la transcription de parole, des actes de dialogue, de gestes des mains et de la tête, et du *focus* d'attention. Une partie du corpus AMI contient des annotations de la cohésion de tâche et de la cohésion sociale réalisée par Hung et. al., [30]. Les réunions ont été divisées en segments de deux minutes. 100 segments ont été sélectionnés au sein de 10 réunions où les équipes de travail ont pour tâche de concevoir une télécommande, et 20 autres segments au sein de deux groupes impliqués dans de vraies discussions. Les données ont été annotées manuellement par 21 annotateurs, utilisant un questionnaire de 27 questions avec une échelle de Likert à 7 points. Le questionnaire utilisé pour recueillir les annotations est une combinaison de plusieurs études état de l'art en psychologie[1]. Chaque segment a été annoté par trois annotateurs différents. L'accord inter-annotateurs a été mesuré grâce au kappa. Au total, 61 segments avec un kappa supérieur à 0.3 ont été retenus. Cela représente 50 segments avec un score de cohésion élevé et 11 segments avec un score faible. Parmi les 61 segments annotés, seulement 25 contiennent des annotations des actes de dialogue. Plus précisément, ces annotations ne sont disponibles que pour 8 des 11 segments avec un score élevé. Par conséquent, notre travail ne considère que 16 segments : 8 avec un score élevé ($M$= 2.995, $SD$= 0.3276) et 8 avec un score faible ($M$=5.994, $SD$= 0.1929).

[1]D'autres questionnaires ont également développés pour annoter la cohésion, par exemple Carron et ses collègues [16] ont développés un questionnaire, néanmoins celui-ci n'était pas adapté à notre objectif puisqu'il permet d'évaluer la cohésion dans des équpes de sport

*Indices sociaux non-verbaux.* Avec l'outil ELAN, nous avons annoté manuellement le *focus* d'attention i. e., *gaze behaviour* de chaque participant du groupe. Nous avons défini 4 *gaze target* pour un participant donné i. e., les trois autres participants du groupe et la classe *autre* (par exemple regard vers la table, les diapos, etc. Le ***Mutual Gaze*** est calculé sur la base d'un échange de regard entre n'importe lequel des participants à un moment donné de l'interaction. La durée du ***Overall Gaze*** est calculée selon le temps total passé par chaque participant dans le groupe à regarder les autres participants. Nous avons également annoté les hochements de tête i. e., les mouvement verticaux de bas en haut. Nous avons utilisé OpenFace [2] pour extraire automatiquement les ***action units*** faciales. Cet outil offre deux type de scores pour les AU, l'intensité et la présence. Le premier indique l'intensité sur une échelle continue de 1 à 5. Le second indique la présence ou l'absence. Nous avons segmenté les données vidéos sur la base de l'activation d'une **action unit** donnée et nous avons calculé la durée et l'intensité de toutes les ***action units*** activées pour chaque segment. Nous avons extrait les instances de rire à partir des fichiers de transcription disponibles dans le corpus. Le tableau 1 montre le nombre d'instances annotées pour l'ensemble de 16 segments. Pour chacun des indices non-verbaux, nous avons calculé nombre d'instances pour chaque segment, la durée totale, la durée moyenne, et pour les AU l'intensité moyenne.

| Annotation | Cohésion faible | Cohésion élévée |
|---|---|---|
| Mutual Gaze | 202 | 258 |
| Outer Brow Raiser (AU2) | 28 | 26 |
| Brow Lowerer (AU4) | 77 | 59 |
| Lip Corner Puller (AU12) | 52 | 113 |
| Head Nods | 100 | 106 |
| Laughter | 31 | 108 |

**Table 1: Nombre total d'instances annotées pour les 16 segments**

*Actes de dialogue.* 15 catégories d'actes de dialogue (DACT) sont utilisées dans le corpus AMI[2]. Dans le corpus, les DACT sont segmentés selon l'intention exprimée dans un énoncé donné i. e., lorsqu'une nouvelle intention de communication est exprimée, un nouveau segment est créé. Chacune des 15 catégories appartient à une des 4 classes principales. La classe **Minor** (*mineur*) regroupe *Backchannel*, *Stall* (pauses) et *Fragment*. La classe **Task** (*tâche*) concerne les informations échangées et les actions qui peuvent être effectuées en groupe ou individuellement. Elle regroupe les catégories *Inform* (informer), dont l'intention est de donner une information, *Suggest* (suggestion), qui renvoie aux actions d'un autre individu ou du groupe dans son ensemble, et enfin *Assess* (évaluer), qui concerne les commentaires exprimant un jugement. La classe **Elicit** (requête) concerne les requêtes exprimées pour obtenir une information ou compléter une action. Trois catégories sont comprises dans cette classe : *Elicit-Inform*, qui est une requête d'information, *Elicit-Assessment* qui est une requête d'évaluation à propos de ce

[2]Pour la description du schéma d'annotation des actes de dialogue, nous nous sommes appuyés sur le manuel d'annotation disponible via http://groups.inf.ed.ac.uk/ami/corpus/annotation.shtml

qui a été dit ou fait, et *Elicit-Comment-Understanding*. Enfin la classe **Other** (autre) concerne les DACT qui expriment des actes ou des commentaires sociaux à propos de ce qui a été dit précédement. Cette classe comprend *Offer*, une intention liée aux actions propres du locuteur, *Comment-About-Understanding*, un commentaire sur un DACT précédent, *Be-Positive*, dont l'intention est de rendre le groupe plus joyeux, *Be-Negative*, qui exprime un sentiment négatif envers un individu ou le groupe dans son ensemble.

*Interruptions.* Afin d'annoter les interruptions dans le corpus AMI, nous avons défini un schéma d'annotation à trois niveaux qui se fonde sur le schéma proposé par [10]. Le niveau **Communication** concerne l'activité de parole des interlocuteurs : *none* permet de considérer les moments où personne ne parle, *speaker*, un seul locuteur parle, *both* deux locuteurs parlent simultanément, *multi*, plus de deux locuteurs parlent simultanément. Le niveau **Transition** concerne les événements de transitions entre le silence et la parole, et inversement, pour le même locuteur ou entre plusieurs locuteurs. *Pause within* est un long silence au sein d'un tour de parole sans changement de locuteur; *Pause between* marque un changement entre le locuteur courant et un autre participant, et inversement, entre un silence entre les deux; *Perfect* est un changement de tour de parole sans silence ou chevauchement; *Overlap within* est un chevauchement sans changement de locuteur. Ce niveau fait également une distinction entre chevauchement et **backchannel** en utilisant les annotations disponibles de actes de dialogue. Le niveau **Interruption** définit les types d'interruption en se fondant sur les données temporelles de l'interruption. Il comprend *overlapped interruption* – une interruption avec un chevauchement et avec le changement de locuteur ; and *paused interruption* – une interruption avec un changement de locuteur et avec un silence où le locuteur ne peut pas terminer son énoncé. A ce niveau, ont également été annotées les occurrences d'interruption, où la personne qui interrompt s'adresse à quelqu'un qui n'est pas la personne interrompue. Par exemple, le locuteur *A* est interrompu par *B* qui s'adresse à *C* [43]. Nous appelons ce type d'interruption *interruption-other* en opposition avec *interruption* où *B* interrompt *A* pour s'adresser à *A*.

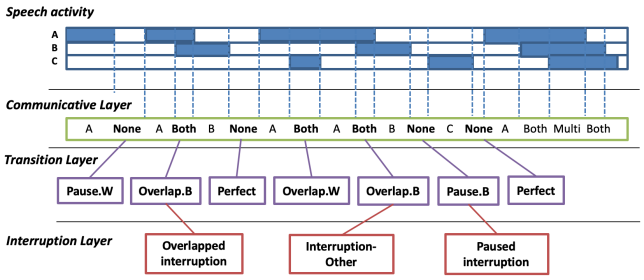



**Figure 1: Exemple d'annotations aux niveaux Communication, Transition et Interruption [10]**

Pour annoter les données, nous avons réalisé une annotation semi-automatique des niveaux **Communication** et **Transition** en prenant en compte le temps de début et de fin de chaque énoncé ainsi que les informations sur les actes de dialogue. Ensuite, nous avons annoté manuellement les interruptions avec l'aide des informations multimodales i. e., parole, transcriptions, le *focus* visuel d'attention i. e., direction de regard du locuteur.

# 4 ANALYSE DE LA COHÉSION NIVEAU PAR NIVEAU

## 4.1 Cohésion et indices sociaux non-verbaux

Pour vérifier nos hypothèses, nous avons effectué un t-test sur les données. Tout d'abord, nous avons vérifié le caractère normal de la distribution des données en utilisant le test de Shapiro-Wilk. Pour les données non-normales, nous avons utilisé le test de Mann-Whitney.

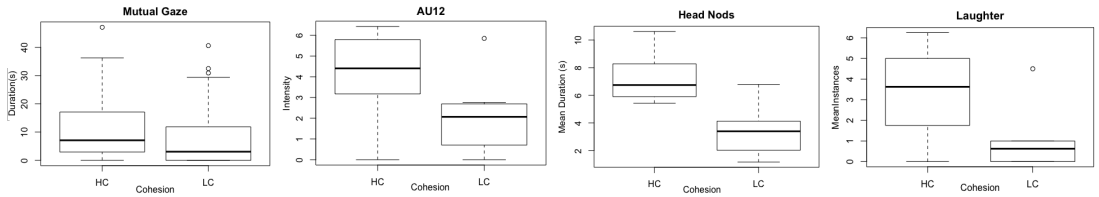
*Gaze.* Au niveau du segment, nous n'avons pas trouvé de différence significative concernant le **gaze** entre les segments ayant un score de cohésion faible et ceux ayant un score élevé ($p$ <0.1). Par conséquent, nous avons analysé le **gaze** au niveau des participants. La durée du regard pour tout participant donné était significativement plus haute entre les participants ($t(64)$ = -2.67, $df$ = 60.75, $p$ = .006) dans les segments ayant un score de cohésion élevé ($M$ = 76.64, $SD$ = 27.83) que dans les segments avec un score faible ($M$ = 59.25, $SD$ = 24.09). De manière similaire, les participants échangent des regards plus longs dans les segments avec un score de cohésion élevé que dans ceux dont le score est faible($U$ = 857, $p$ = .03, $r$ = .31).

*Facial Action Units.* Sur la base de nos annotations, nous avons observé que l'AU12, i. e., haussement du coin de la lèvre, était activée plus fréquemment dans le groupe montrant davantage de cohésion. La durée de l'activation est significative plus haute ($t(16)$ = -2.57, $df$ = 10.35, $p$= .026) dans les segments avec un score de cohésion élevé ($M$ = 65.05, $SD$ = 42.25) que dans les segments avec un score faible ($M$ = 21.91, $SD$ = 21.34). De plus, l'intensité moyenne de l'AU12 lorsqu'elle est activée est également plus élevée mais la différence n'apparaît pas être significative ($t(16)$ = -2.04, $df$ = 13.77, $p$ = .060). Il n'y a pas de différence significative au niveau de la durée ou de l'intensité d'activation de l'AU2 i. e., **haussement extérieur du sourcil** et de l'AU4 i. e., **haussement des sourcils**.

*Head Nods.* Même si il n'y a pas de différence très importante concernant le nombre d'occurrences de *head nods* pour les deux groupes, il y a néanmoins une différence significative au niveau de la durée ($t(16)$ = -4.33, $df$ = 13.99, $p$ = .0006). Généralement, les *head nods* durent plus longtemps dans les segments avec un score de cohésion élevé ($M$= 7.23, $SD$ = 3.09) que dans ceux avec un score faible ($M$ = 3.38, $SD$ = 3.23).

*Rire.* Le rire est plus fréquent dans les segments avec un score de cohésion élevé. La durée n'est pas significativement différente mais le nombre moyen d'occurrence est plus bas ($t(16)$ = -2.59, $df$ = 12.45, $p$ = .022) dans les segments avec un faible score ($M$ = 0.96, $SD$ = 2.22) que dans ceux avec un score élevé ($M$ = 3.37, $SD$ = 4.64).

*Discussion.* Comme expliqué dans la section 2.3, notre objectif est d'identifier les indices sociaux non-verbaux qui sont associés à des scores élevés ou faible de cohésion. Pour cela, nous avons observé les comportement tels que le *gaze*, les *action units*, les *head nods* et le rire. Notre hypothèse était que des indices associés à des affects positif, à l'implication et au soutien e. g., le fait fixer du regard un locuteur, le rire, les hochements de tête, seraient plus important au sein de groupe montrant une forte cohésion. Nos résultats montrent




**(a) Durée moyenne du *Mutual Gaze* ($p$ = .030)** **(b) Intensité moyenne de l'AU12 ($p$ = .026)** **(c) Durée moyenne des *Head Nods* ($p$ = .0006)** **(d) Moyenne des instances de Rire ($p$ = .022)**

**Figure 2: Valeurs moyennes des indices sociaux non-verbaux pour les segments avec une cohésion faible et élevée**

que les instances de rire sont très importantes dans les groupes avec une forte cohésion. Nous avons observé que les instances où plus d'un participant partage un rire sont plus importantes. Cela confirme ce que plusieurs études sur le rire dans les groupes ont montré : "Le rire établit une forme de lien entre les membres d'un groupe et rend les participants plus à l'aise" [26]. En moyenne, un rire dure 7 secondes dans les segments avec une cohésion élevée. De plus, nous observons que l'AU12, qui est associée à la joie et au sourire [19], a une intensité plus élevée dans ce type de segments. De même, nous avons observé que l'AU4, qui est souvent associée à la colère et au mépris [46], apparaît plus souvent dans les segments avec une cohésion faible. Néanmoins, la différence n'est pas significative. Cela peut être dû au fait que nous n'avons observé que des segments relativement courts en terme de durée (2min) et peut-être qu'en considérant davantage de segments dans le corpus, une différence significative pourrait apparaître. Ensuite, concernant les *head nods*, ces derniers créent souvent un environnement favorable [28] et ils sont communément associés à une écoute attentive. Dans nos données, il n'y a presque aucune différence dans la fréquence d'occurrences des hochements de tête entre les deux groupes (cohésion élevée vs faible). Néanmoins, nous avons observé une différence significative au niveau de la durée moyenne. Enfin, concernant le **gaze**, nous supposions que dans le groupe avec une forte cohésion, les participants passeraient plus de temps à fixer les autres du regard ou à maintenir un échange de regard. Les résultats montrent que dans les segments avec un score élevé de cohésion les participants fixent du regard plus longtemps que dans les autres segments. Ces résultats vont dans le sens des études analysant le *eye-gazing* comme un régulateur de la compréhension et un moyen de gérer flux dans les interactions multipartites [32]. De plus, le groupe avec un cohésion faible passe moins de temps à soutenir le regard des autres participants. Exline et. al., [21] expliquent en effet que la durée de contact visuel décroît dans des interactions non-collaboratives.

## 4.2 Cohesion et actes de dialogue

Le nombre moyen de DACT par segment dans notre *dataset* est de 52. Ce sont les DACT de la classe *Task* qui sont le plus représentés par segment (54, pour le segment en contenant le plus et 18 pour celui en contenant le moins). *A contrario*, *Other* a le nombre le plus bas par segment (6 pour celui en contenant le plus, 0 pour celui en contenant le moins). Afin de comprendre en quoi les DACT peuvent être liés à la cohésion, nous avons vérifié si le nombre de DACT de chaque catégorie avait un impact sur le score de cohésion, en supposant ainsi que certaines catégories seraient positivement corrélées et d'autres négativement corrélées au score de cohésion.

*Les quatre classes principales de DACT.* Pour chacun des 16 segments, nous avons d'abord mesuré la corrélation entre le nombre de DACT de chacune des quatre classes principales (*Task*, *Elicit*, *Minor* and *Other*) et le score de cohésion. Nous avons considéré comme variable indépendante le nombre de DACT pour chaque catégorie et comme variable dépendante le score de cohésion. Pour mesurer la corrélation entre les deux variables, nous avons utilisé le test de Pearson. Nous n'avons trouvé aucune corrélation entre les deux variables, la valeur de $p$ étant supérieure à .05, les résultats ne sont pas significatifs et le coefficient de corrélation ne peut pas être interprété.

*Les 15 catégories spécifiques de DACT.* Puisque aucune corrélation entre le score de cohésion et le nombre de DACT n'a pu être établie pour aucune des quatre classes principales, nous avons considéré que ces classes étaient trop larges pour montrer des résultats significatifs et qu'une corrélation existait peut-être entre le score de cohésion et le nombre de DACT de chaque catégories spécifiques. Nous avons utilisé un test de Pearson pour mesurer la corrélation. Pour la plupart de catégories, les résultats ne sont pas significatifs ($p$ est inférieur à .05). Une seule catégorie, *Be-Positive* montre un résultat significatif ($p$ = .030). Le coefficient de corrélation est supérieur à 0.5, ce qui montre une corrélation élevée entre les DACT *Be-Positive* et le niveau de cohésion. Ces résultats soutiennent l'hypothèse posée par [36] concernant l'impact des expressions de sentiments sur le niveau de cohésion.

*Régression linéaire avec contraste entre les 4 classes principales.* Pour vérifier les résultats obtenus avec le test de corrélation de Pearson, nous avons effectué une régression linéaire avec des contrastes entre les quatre classes[3]. Ce test montre une différence entre le score moyen de cohésion obtenu avec une classe en contraste avec celui obtenu avec les trois autres. Les résultats confirment ceux obtenus avec le Pearson. Lorsque l'on met en contraste chacune de quatre classes, avec les trois autres, aucune ne montre un impact significatif sur le score de cohésion (l'écart entre le score moyen n'est jamais supérieur à 0.1 ou inférieur à -0.1)

*Discussion.* A l'exception des actes *Be-positive*, notre analyse ne montre aucune corrélation entre le nombre d'actes de dialogue apparaissant dans un segment et le score de cohésion. Ces résultats peuvent être expliqués par la structure de la conversation. Puisque

[3]A cause du nombre important de sous-catégories (15), nous n'avons pu mesurer les contrastes qu'entre les 4 classes principales

| DACT | Coef. Corr | p-value |
|---|---|---|
| Inform | -0.485 | .056 |
| Suggest | 0.373 | .154 |
| Assess | 0.452 | .078 |
| Elicit-Inform | -0.194 | .470 |
| Elicit-Offer-or-Suggestion | 0.373 | .155 |
| Elicit-Comment-Understanding | 0.237 | .377 |
| Elicit-Assess | -0.097 | .721 |
| Offer | -0.388 | .138 |
| Comment-About-Understanding | -0.316 | .232 |
| Be-Positive | 0.542 | .030 |

**Table 2: Coefficients de corrélation et p-values pour le test de Pearson entre le score de cohésion et le nombre de DACT de chaque catégorie spécifique dans chaque segment**

les interactions sont orientées tâche – l'objectif de ces réunions est d'organiser le travail en équipe – les changements de locuteur sont très fréquents (33 fois en moyenne par segment). Les locuteurs lorsqu'ils prennent la parole ne produisent pas nécessairement un DACT qui peut former une paire adjacente avec le tour précédent (nous avons estimé que seulement la moitié d'entre eux forment une paire adjacente). Ce type de structure peut rendre difficile pour l'annotateur de prendre en compte les actes de dialogue pour annoter le niveau de cohésion. Une prochaine étude devrait se concentrer sur les mécanismes d'ancrage et de synergie comme décrit dans [18]. Une autre hypothèse pour expliquer ces résultats est que les actes de dialogue sont peut-être liés à la cohésion mais uniquement lorsqu'on les prend en compte conjointement à d'autres indices multimodaux. C'est ce que nous explorons dans la section 5).

## 4.3 Cohésion et Interruptions

Notre objectif est d'analyser la relation entre le *turn taking*, les interruptions et la cohésion. Nous avons utilisé le test de Pearson pour mesurer la corrélation entre les segments et les variables indépendantes et avons réalisé une analyse de la variance pour mesurer les différences entre les deux groupes.

*Tours de parole.* Le nombre de tours est positivement corrélé au score de cohésion ($r$ = 0.624, $p$ = .01). L'analyse de la variance (one-way ANOVA) montre qu'il y a un effet du score de cohésion sur le nombre de tours de parole pendant l'interaction ($f(1, 14)$ = 6.465, $p$ = .023). Les segments avec un score de cohésion élevé montrent que les changements de tours sont plus fréquents ($M$ = 23.75, $SD$ = 7.741) que dans les groupes avec un score faible ($M$ = 15.125, $SD$ = 5.667).

*Overlaps.* Le nombre de chevauchements est positivement corrélé à la cohésion ($r$ = 0.519, $p$ = .039). Le nombre de chevauchements est significativement plus important dans le groupe avec un score de cohésion élevé ($M$ = 27.62, $SD$ = 4.92) que dans les groupes avec un score faible ($M$ = 16.5, $SD$ = 11.46), avec ($F(1, 14)$ = 5.327, $p$ = .037).

*Overlapped Interruptions.* De même, on constate une corrélation positive entre le nombre de *overlapped interruptions* et le score de cohesion ($r$ = 0.613, $p$ = .008). L'analyse de la variance (one-way ANOVA) montre une différence significative du nombre de *overlapped interruptions* dans les segments avec un score élevé et ceux avec un score faible ($F(1, 14)$ = 9.847, $p$ = .007). Les segments avec un score élevé montre un nombre plus important d'interruptions ($M$ = 9.75, $SD$ = 3.327) en comparaison à ceux ayant un score faible ($M$ = 4.62, $SD$ = 3.20). En revanche, aucune corrélation entre la cohésion et les *paused interruptions*. Le t-test montre que les scores sont significativement plus hauts pour les *overlapped interruptions* ($M$ = 7.187, $SD$ = 4.118) que pour les *paused interruptions* ($M$ = 2.937, $SD$ = 2.205).

*Interruptions-other.* Les occurrences de ces interruptions sont en corrélation positive avec la cohésion ($r$ = 0.674, $p$ = .004). L'analyse de la variance (one-way ANOVA) montre que ($F(1, 14)$ = 0.994, $p$ = .007) les participants utilisent un nombre plus important d'*interruption-other* dans les segments avec un score élevé ($M$ = 2.125, $SD$ = 1.124) que dans ceux avec un score faible ($M$ = 0.50, $SD$ = 0.756). Le t-test montre que les scores sont significativement plus élevés pour les *overlapped interruptions* ($M$ = 7.187, $SD$ = 4.11) que pour *interruption-other* ($M$ = 1.31, $SD$ = 1.30), ($t(16)$ = 7.388, $p$ <.01).

| Tours et Interruptions | Coef. Correl. | p-value |
|---|---|---|
| Tours de parole | 0.624 | .010 |
| Overlaps | 0.519 | .039 |
| Overlapped interruption | 0.613 | .008 |
| Paused interruption | 0.258 | .334 |
| Interruption-other | 0.674 | .004 |

**Table 3: Coefficients de corrélation et p-values de test Pearson entre la cohésion et les caractéristiques au *turn taking* et aux interruptions**

*Discussion.* Le tableau 3 résume la corrélation entre cohésion et les caractéristiques liées aux *turn taking* et aux interruptions. Notre hypothèse que les tours de parole sont plus importants dans le groupes montrant une cohésion forte et plus limités dans le groupe où la cohésion est plus faible est validée. Ces résultats confirment ceux de Hung et. al., [30]. Les résultats montrent que les participants échangent et prennent la parole plus fréquemment dans les groupes où la cohésion est forte que dans ceux où elle est plus faible. Cela s'explique par le fait que les membres du groupe participent activement à l'interaction, ce qui fait augmenter le nombre de tours. Cela fait également baisser le temps entre deux tours de parole successifs, par rapport aux groupes avec un score faible.

Les occurrences de chevauchement durant les interactions sont positivement corrélées au score de cohésion. Nous pensons que cela est dû à la nature du scénario sur lequel se fonde le corpus AMI : des réunions orientées tâche où les participants doivent collaborer et discuter pour accomplir un objectif commun. Notre hypothèse concernant le nombre d'interruptions dans les segments avec un score de cohésion élevé est donc validée. Ce résultat est en accord avec ceux de Tannen [45], qui décrit les interruptions comme de bons indicateurs de cohésion dans les groupes, par exemple lorsque les participants sont en mesure de compléter les énoncés des uns des autres.

Les résultats concernant *interruption-other* confirment également ce qu'affirment les études en psychologie [3, 39] concernant

les occurrences de ces interruptions. Les résultats montrent également que le nombre de ces interruptions est relativement faible en comparaison au nombre de *overlapped interruptions*. Néanmoins, pour concevoir un modèle de conversation multimodal pour les interaction multipartites les *interruptions-other* ne peuvent pas être ignorées.

## 5 ANALYSE MULTI-NIVEAU DE LA COHÉSION

Dans la section précédente, notre analyse considérait trois niveaux spécifiques de manière distinctes (le verbal, les indices sociaux non-verbaux et les interruptions) afin de vérifier leur impact individuel sur le score de cohésion. Les indices sociaux non-verbaux comme le *mutual gaze*, le rire et l'AU12 sont associés à des scores de cohésion élevés. Concernant les actes de dialogue, les résultats montrent que le nombre d'occurrences d'acte de dialogue appartenant à des catégories spécifiques n'ont pas d'impact sur le niveau de cohésion, à l'exception des actes de type *Be-Positive* qui apparaissent comme étant positivement corrélés à la cohésion. Le nombre de tours de parole, d'interruptions et de chevauchements est positivement corrélés à la cohésion. L'analyse de ces trois niveaux montre que la perception de la cohésion s'appuie sur des comportements appartenant à différentes modalités. Néanmoins, pour une analyse multimodale de la cohésion de groupe, nous avons besoin d'analyser comment ces comportements co-occurrents et quel impact cette co-occurrence a sur le niveau de cohésion. Sur la base de la littérature existante, nous avons analysé la relation entre différents comportements : (i) interruptions – *gaze* et cohésion (ii) actes de dialogue – *head nods* et cohésion.

*Interruptions et Gaze.* Le *eye gaze* aide significativement la prédiction de la prise de parole [31]. Les résultats de la section 4.1 montrent que les paires de participants soutiennent plus longtemps le regard de l'un et l'autre dans les groupes avec un score de cohésion élevé. Les instances de *mutual gaze* apparaissant pendant les interruptions sont positivement corrélées à la cohésion (le test de Pearson donne $r = 0.731$, $p = .001$). L'analyse de la variation montre une différence significative entre le nombre d'instances de *mutual gaze* ($F(1,14)=15.868$, $p = .001$) i. e., le nombre de *mutual gaze* est plus important dans les groupes comme une cohésion élevée ($M = 4.875$, $SD = 2.167$) que dans ceux avec une cohésion faible ($M = 1.25$, $SD = 1.388$). Les participants pendant les interruptions se regardent l'un l'autre plus fréquemment dans les groupes avec une cohésion forte que dans ceux avec une cohésion faible.

*Actes de dialogue et head nods.* Toute conversation met en jeu à la fois des signaux verbaux et des signaux non-verbaux. Dans cette section, nous présentons l'analyse de la co-occurrence des *head nods* et des DACT en rapport avec le score de cohésion. Pour cela, nous avons extrait les instances de *head nods* effectués par les participants en train d'écouter et les types d'acte de dialogue exprimés simultanément par le participant ayant la parole. Nous avons ensuite effectué une régression linéaire avec contrastes entre les quatre classes principales de DACT. Le premier modèle établit un contraste entre la classe *task* et les trois encore lorsqu'elles sont produites simultanément avec un *head nod*. Le score moyen de cohésion obtenu quand ces DACT apparaissent avec un *head nod* est de 4.955. Les résultats montrent qu'un *head nod* est effectué alors que le locuteur exprime un DACT de type *task* celui-ci est alors associé à un score de cohésion plus bas que lorsqu'il est en co-occurrence avec l'un des trois autres DACT (−0.200 ). Dans le même modèle, le contraste résiduel entre la classe *elicit* et *other*, lorsqu'elles co-occurrent avec un *head nod* montre que la classe *elicit* est associée à un score plus haut que *other* (1.042). Le second modèle met en contraste *other* avec *task*, *elicit* et *minor* lorsqu'ils co-occurrent avec un head nod. Les résultats montrent que les DACT de la classe *other* produisent un niveau de cohésion plus bas que les trois autres lorsqu'ils sont exprimés simultanément à avec un *head nod* (−0.298). Dans le même modèle, le contraste résiduel entre *task* et *elicit* montre que *task* produit un score de cohésion plus bas que *elicit* dans ce type de contexte.

A un niveau multimodal, i. e., interruptions – *gaze* et actes de dialogue – *head nods*, nous observons que certains comportements qui ne sont pas statistiquement associés au score de cohésion lorsqu'ils sont pris individuellement, le sont en revanche lorsqu'ils sont envisagés conjointement. A partir de cette analyse, nous pouvons conclure que les comportements multimodaux sont des indices du niveau de cohésion, et de fait peuvent aider à son estimation.

## 6 CONCLUSION ET TRAVAUX FUTURS

Dans cet article, nous proposons une analyse de la cohésion, dans les interactions multipartites, qui se concentrent sur trois niveaux d'analyse i. e., *les indices sociaux non-verbaux*, *les actes de dialogue* et *les interruptions*. Lorsqu'ils sont envisagés séparément les interruptions et certains indices sociaux non-verbaux ont un impact sur le niveau de cohésion. Cet article montre également qu'il est essentiel de combiner différentes modalités pour procéder à une analyse pertinente de la cohésion. Les résultats de cette étude vont contribuer à développer un modèle computationnel pour simuler un groupe d'agents virtuels avec cohésion. Dans une prochaine étude, nous avons pour objectif de répliquer ces résultats sur un autre corpus d'interactions multipartites mais aussi de développer un modèle d'estimation automatique de la cohésion.

*Remerciements.* Ce projet est financé par le programme de recherche et d'innovation Horizon 2020 de l'Union Européenne sous le numéro d'agrément 769553. Ces résultats ne sont le reflet que du point de vue des auteurs et n'impliquent aucunement la responsabilité de l'Union Européenne quant à l'utilisation qu'il peut être faite des informations qu'ils contiennent. Nous souhaitons adresser nos remerciement au Dr. Hayley Hung de TUDelft pour avoir partagé ses données sur la cohésion. Nous souhaitons également remercier Céline Caristan et Amandine Guillin pour leur aide sur les analyses statistiques.

## REFERENCES

[1] K. Back. 1951. Influence through social communication. *The Journal of Abnormal and Social Psychology* 46, 1 (1951), 9.
[2] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 2016. Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 1–10.
[3] Adrian Bangerter, Eric Chevalley, and Sylvie Derouwaux. 2010. Managing third-party interruptions in conversations: Effects of duration and conversational role. *Journal of Language and Social Psychology* 29, 2 (2010), 235–244.
[4] D. Beal, R. Cohen, M. Burke, and C. McLendon. 2003. Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of applied psychology* 88, 6 (2003), 989.

[5] Geoffrey W Beattie. 1981. Interruption in conversational interaction, and its relation to the sex and status of the interactants. *Linguistics* 19, 1-2 (1981), 15–36.

[6] D. Bohus and E. Horvitz. 2010. *Computational Models for Multiparty Turn Taking.* Technical Report. Microsoft Research Technical Report MSR-TR 2010-115.

[7] L. Braaten. 1991. Group cohesion: A new multidimensional model. *Group* 15, 1 (1991), 39–55.

[8] Harry Bunt. 2011. The semantics of dialogue acts. In *Proceedings of the Ninth International Conference on Computational Semantics*. Association for Computational Linguistics, 1–13.

[9] Angelo Cafaro, Nadine Glas, and Catherine Pelachaud. 2016. The effects of interrupting behavior on interpersonal attitude and engagement in dyadic interactions. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 911–920.

[10] A. Cafaro, B. Ravenet, and C. Pelachaud. 2019. Exploiting evolutionary algorithms to model nonverbal reactions to conversational interruptions in user-agent interactions. *IEEE Transactions on Affective Computing* (2019), 1–1. https://doi.org/10.1109/TAFFC.2019.2947054

[11] Sabrina Campano, Caroline Langlet, Nadine Glas, Chloé Clavel, and Catherine Pelachaud. 2015. An ECA expressing appreciations. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 962–967.

[12] S. Carless and C. De Paola. 2000. The measurement of cohesion in work teams. *Small group research* 31, 1 (2000), 71–88.

[13] J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, et al. 2005. The AMI meeting corpus: A pre-announcement. In *International workshop on machine learning for multimodal interaction*. Springer, 28–39.

[14] A. Carron. 1982. Cohesiveness in Sport Groups: Interpretations and Considerations. *Journal of Sport psychology* 4, 2 (1982), 123–138.

[15] A. Carron and P. Chelladurai. 1981. The dynamics of group cohesion in sport. *Journal of Sport Psychology* 3, 2 (1981), 123–139.

[16] A. Carron, N. Widmeyer, and L. Brawley. 1985. The development of an instrument to assess cohesion in sport teams: The Group Environment Questionnaire. *Journal of sport psychology* 7, 3 (1985), 244–266.

[17] M. Casey-Campbell and M. Martens. 2009. Sticking it all together: A critical assessment of the group cohesion–performance literature. *International Journal of Management Reviews* 11, 2 (2009), 223–246.

[18] Pierre Dillenbourg and David Traum. 2006. Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *The Journal of the Learning Sciences* 15, 1 (2006), 121–151.

[19] Paul Ekman, Richard J Davidson, and Wallace V Friesen. 1990. The Duchenne smile: Emotional expression and brain physiology: II. *Journal of personality and social psychology* 58, 2 (1990), 342.

[20] Rosenberg Ekman. 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA.

[21] Ralph V Exline. 1963. Explorations in the process of person perception: Visual interaction in relation to competition, sex, and need for affiliation. *Journal of personality* (1963).

[22] S. Fang and C. Achard. 2018. Estimation of Cohesion with Feature Categorization on Small Scale Groups. (2018). WACAI.

[23] L Festinger, S. Schachter, and K. Back. 1950. Social pressures in informal groups; a study of human factors in housing. (1950).

[24] D. Forsyth. 2018. *Group dynamics*. Cengage Learning.

[25] Riccardo Fusaroli, Joanna Rączaszek-Leonardi, and Kristian Tylén. 2014. Dialog as interpersonal synergy. *New Ideas in Psychology* 32 (2014), 147–157.

[26] Phillip Glenn. 2003. *Laughter in interaction*. Vol. 18. Cambridge University Press.

[27] P. Goodman, E. Ravlin, and M. Schminke. 1987. Understanding groups in organizations. *Research in Organizational Behavior* (1987).

[28] Uri Hadar, Timothy J Steiner, Ewan C Grant, and F Clifford Rose. 1984. The timing of shifts of head postures during conservation. *Human Movement Science* 3, 3 (1984), 237–245.

[29] Mattias Heldner and Jens Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics* 38, 4 (2010), 555 – 568.

[30] H. Hung and D. Gatica-Perez. 2010. Estimating cohesion in small groups using audio-visual nonverbal behavior. *IEEE Transactions on Multimedia* 12, 6 (2010), 563–575.

[31] Kristiina Jokinen, Hirohisa Furukawa, Masafumi Nishida, and Seiichi Yamamoto. 2013. Gaze and Turn-taking Behavior in Casual Conversational Interactions. *ACM Trans. Interact. Intell. Syst.* 3, 2, Article 12 (Aug. 2013), 30 pages.

[32] Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta psychologica* 26 (1967), 22–63.

[33] Han Z Li. 2001. Cooperative and intrusive interruptions in inter-and intracultural dyadic discourse. *Journal of Language and Social Psychology* 20, 3 (2001), 259–284.

[34] Gregory J Mills. 2014. Dialogue in joint activity: Complementarity, convergence and conventionalization. *New ideas in psychology* 32 (2014), 158–173.

[35] P. Mudrack. 1989. Defining group cohesiveness: A legacy of confusion? *Small Group Behavior* 20, 1 (1989), 37–49.

[36] Marjolein C Nanninga, Yanxia Zhang, Nale Lehmann-Willenbrock, Zoltán Szlávik, and Hayley Hung. 2017. Estimating verbal expressions of task and social cohesion in meetings by quantifying paralinguistic mimicry. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. ACM, 206–215.

[37] Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences* 27, 2 (2004), 169–190.

[38] W. Piper, M. Marrache, R. Lacroix, A. Richardsen, and B. Jones. 1983. Cohesion as a basic bond in groups. *Human Relations* 36, 2 (1983), 93–108.

[39] Clotilde Pontecorvo, Sabine Pirchio, and Laura Sterponi. 2000. Are there just two people in a dyad? Dyadic configurations in multiparty family conversations. *Schweizerische Zeitschrift für Bildungswissenschaften* 22, 3 (2000), 535–558.

[40] David Reitter, Frank Keller, and Johanna D Moore. 2006. Computational modelling of structural priming in dialogue. In *Proceedings of the human language technology conference of the naacl, companion volume: Short papers*. Association for Computational Linguistics, 121–124.

[41] Virginia P Richmond, James C McCroskey, and Steven K Payne. 1991. *Nonverbal behavior in interpersonal relations*. Prentice Hall Englewood Cliffs, NJ.

[42] J. Santoro, A. Dixon, C. Chang, and S. Kozlowski. 2015. Measuring and monitoring the dynamics of team cohesion: Methods, emerging tools, and advanced technologies. In *Team cohesion: Advances in psychological theory, methods and practice*. Emerald Group Publishing Limited, 115–145.

[43] E. Schegloff. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in society* 29, 1 (2000), 1–63.

[44] María Teresa Taboada. 2004. *Building coherence and cohesion: Task-oriented dialogue in English and Spanish*. Vol. 129. John Benjamins Publishing.

[45] Deborah Tannen. 1994. *Gender and discourse*. Oxford University Press.

[46] Y-I Tian, Takeo Kanade, and Jeffrey F Cohn. 2001. Recognizing action units for facial expression analysis. *IEEE Transactions on pattern analysis and machine intelligence* 23, 2 (2001), 97–115.

[47] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. 2009. Social signal processing: Survey of an emerging domain. *Image and vision computing* 27, 12 (2009), 1743–1759.

[48] Wen Wang, Kristin Precoda, Raia Hadsell, Zsolt Kira, Colleen Richey, and Gabriel Jiva. 2012. Detecting leadership and cohesion in spoken interactions. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5105–5108.

[49] Candace West and Don H Zimmerman. 2015. Small insults: A study of interruptions in cross-sex conversations between unacquainted persons.. In *American Sociological Association's Annual Meetings, Sep, 1978, San Francisco, CA, US;*.

[50] Y. Zhang, J. Olenick, C. Chang, S. Kozlowski, and H. Hung. 2018. TeamSense: Assessing Personal Affect and Group Cohesion in Small Teams through Dyadic Interaction and Behavior Analysis with Wearable Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 150.