



UNIVERSIDADE DE SÃO PAULO

Escola de Artes, Ciências e Humanidades

Relatório Técnico PPgSI-002/2023
*Manual de Anotação do corpus Porttinari-base
com Atos de Fala*

Versão 1.0

Nataly Leopoldina Patti da Silva
Norton Trevisan Roman
Ariani Di Felippo

Novembro - 2023

O conteúdo do presente relatório é de única responsabilidade dos autores.

Série de Relatórios Técnicos

PPgSI-EACH-USP

Rua Arlindo Béttio, 1000 – Ermelino Matarazzo

03828-000 – São Paulo, SP.

TEL: (11) 3091-8197

<http://www.each.usp.br/ppgsi>

Manual de Anotação do corpus Porttinari-base com Atos de Fala

Versão 1.0

Nataly Leopoldina Patti da Silva¹
Norton Trevisan Roman¹
Ariani Di Felippo²

¹Escola de Artes, Ciências e Humanidades – Universidade de São Paulo
São Paulo – SP, Brasil

² Núcleo Interinstitucional de Linguística Computacional (NILC)
Departamento de Letras, Universidade Federal de São Carlos
São Carlos – SP, Brasil

natalypatti@usp.br

norton@usp.br

arianidf@gmail.com

Resumo. *A teoria dos Atos de Fala estuda como as palavras trazem não apenas informações e significados, mas também expressam ações [Searle 1969], sendo desta forma essencial para a compreensão da intenção comunicativa pretendida no ato da fala. Este documento apresenta o manual de anotação do corpus Porttinari-base [Lopes et al. 2022] com atos de fala seguindo o padrão proposto pelo ISO-24617-2 [ISO 2012].*

1. Introdução

O presente documento apresenta o manual de anotação do corpus Porttinari-base [Lopes et al. 2022] com atos de fala seguindo o padrão proposto pelo ISO-24617-2 [ISO 2012]. A teoria dos Atos de Fala estuda como as palavras, organizadas em orações e sentenças, trazem não apenas informações e significados, mas também expressam ações [Searle 1969]. O ato de fala representa a intenção comunicativa pretendida no ato da fala. Ou seja, através da língua é possível realizar uma ação ou levar alguém a realizar uma ação, como agradecer, questionar, pedir, prometer etc. [Austin 1962].

A anotação de atos de fala consiste na associação de classes de atos de fala (perguntar, pedir, afirmar, prometer etc.) a cada enunciado em determinados contextos, como falas em um diálogo, tweets ou frases de um texto, com o objetivo de identificar a função comunicativa realizada por aquele enunciado. No caso do presente trabalho, a tarefa consiste na associação de uma classe de ato de fala da taxonomia do ISO-24617-2 a cada sentença do corpus Porttinari.

O corpus Porttinari (“PORTuguese Treebank”) compreende um conjunto de textos jornalísticos na língua portuguesa, oriundos do jornal Folha de São Paulo [Folha 1921], anotado manualmente com etiquetas morfossintáticas de acordo com o modelo de Dependências Universais (UD) [Nivre et al. 2020]. O corpus faz parte do projeto PORTuguese processing - Towards Syntactic Analysis and parsing (POeTiSA)¹, que é um trabalho de longo prazo que visa a alcançar resultados de ponta em recursos baseados em sintaxe e

¹<https://sites.google.com/icmc.usp.br/poetisa/>

no desenvolvimento de aplicações voltadas para a língua portuguesa do Brasil. Na parte dos recursos, o projeto tem como foco a produção de um corpus multigênero baseado em Dependências Universais e textos sintaticamente anotados.

O corpus Porttinari encontra-se em construção, mas é possível encontrar uma versão base do mesmo, Porttinari-base², datada de agosto de 2022. Este corpus corresponde a uma amostra do Porttinari, contendo 8.420 sentenças anotadas manualmente com etiquetas morfológicas e morfo-sintáticas.

Nas próximas seções deste relatório será apresentada uma metodologia para a anotação do corpus com atos de falas, bem como o conjunto de atos de falas, o padrão adotado e um passo a passo para anotação.

2. Metodologia para anotação

Para anotação do corpus foi escolhido o conjunto de atos de fala definido pelo ISO-24617-2. O ISO define, além de um padrão (os conjuntos de dimensões e funções comunicativas) para anotação de diálogos com atos de falas, uma linguagem de anotação, *Dialogue Act Markup Language* (DiAML), e um método de segmentação das falas em unidades básicas. Para os fins de anotação do corpus, foram considerados apenas os conjuntos de dimensões e funções comunicativas definidas pelo padrão.

2.1. Seleção dos rótulos de anotação

O ISO-24617-2 define um conjunto de 56 funções comunicativas divididas em dois grupos e 9 dimensões. A Figura 1 apresenta todas as funções comunicativas do grupo de funções de Propósito Geral e seus respectivos agrupamentos e as funções comunicativas de Propósito Específico e as dimensões que elas estão inseridas. Em negrito estão destacadas apenas as funções comunicativas que foram encontradas durante a anotação do corpus.

Os grupos separam as funções comunicativas entre as que são de Propósito Geral e as que são de Propósito Específico. O grupo de Propósito Geral se refere as funções que podem ser utilizadas com qualquer conteúdo semântico (informação, ação, situação, evento ou objeto), tendo como principal característica obter ou pedir informações e discutir ações. Uma função comunicativa deste grupo pode, por exemplo, informar alguém sobre qualquer tipo de assunto, seja qual for o conteúdo semântico desta informação. Os tipos de funções comunicativas que esse grupo engloba e suas respectivas definições estão apresentadas na Tabela 1.

Por sua vez, as funções de Propósito Específico são divididas em 9 dimensões e são específicas de suas respectivas dimensões, e em sua maioria, não possuem nenhum conteúdo semântico ou apenas um conteúdo semântico marginal. Elas se concentram em algum aspecto específico da comunicação voltado para uma categoria de conteúdo semântico (tipo de informação, ação, situação, evento ou objeto que forma o conteúdo semântico), por exemplo, a dimensão “Turn Management” corresponde à categoria de informação sobre a atribuição da função de orador. Todas as dimensões definidas nesse grupo e suas respectivas definições estão apresentadas na Tabela 2.

²<https://drive.google.com/file/d/1odhlwPFxeGw70jYKuSaGVdIKqcOgu8XP/view>

Funções Comunicativas de Propósito Geral

Information providing functions	Information seeking functions	Comissive functions	Directive functions
<ol style="list-style-type: none"> 1. inform 2. agreement 3. disagreement 4. confirm 5. disconfirm 6. correction 7. answer 	<ol style="list-style-type: none"> 1. question 2. propositional question 3. set question 4. check question 5. choice question 6. test question 	<ol style="list-style-type: none"> 1. promise 2. offer 3. address request 4. accept request 5. decline request 6. address suggest 7. accept suggest 8. decline suggest 	<ol style="list-style-type: none"> 1. instruct 2. request 3. suggest 4. address offer 5. accept offer 6. decline offer

Funções Comunicativas de Propósito Específico

Feedback functions	Turn-management functions	Time-management functions	Social obligations management functions
<ol style="list-style-type: none"> 1. auto positive 2. auto negative 3. allo positive 4. allo negative 5. feedback elicitation 	<ol style="list-style-type: none"> 1. turn accept 2. turn take 3. turn grab 4. turn assign 5. turn release 6. turn keep 	<ol style="list-style-type: none"> 1. stalling 2. pausing 	<ol style="list-style-type: none"> 1. apology 2. thanking 3. compliment 4. congratulation 5. sympathy expression 6. init greeting 7. return greeting 8. init self introduction 9. return self introduction 10. accept apology 11. accept thaking 12. init goodbye 13. return goodbye
Discourse-structuring functions	Own and partner management functions	Contact management functions	
<ol style="list-style-type: none"> 1. interaction structuring 2. opening 3. topic shift 	<ol style="list-style-type: none"> 1. self error 2. retraction 3. self correction 4. completion 5. correct misspeaking 	<ol style="list-style-type: none"> 1. contact check 2. contact indication 	

Figura 1. Lista das funções comunicativas de Propósito Geral e Propósito Específico definidas pelo ISO

Tabela 1. Dimensões de Propósito Geral

Dimensão	Definição
Information Seeking	refere-se a funções comunicativas voltadas a questões de várias formas diferentes, dependendo do tipo de informação que o locutor espera receber.
Information Providing	refere-se aos atos de fala em que o locutor tem o objetivo de informar ou prover informação ao ouvinte. Sua função mais genérica é a “inform”, mas também possui outras mais específicas como “agreement”, “answer”, “disconfirm”, etc.
Commissive	as funções comunicativas dessa dimensão são utilizadas quando o locutor se compromete a executar alguma ação. Alguns exemplos de funções dessa classe são “offer” e “promise”.
Directive	as funções comunicativas dessa dimensão são utilizadas quando o locutor induz o ouvinte a executar alguma ação. Alguns exemplos dessa função são “instruct” e “request”.

Tabela 2. Dimensões de Propósito Específico

Dimensão	Definição
Feedback	as funções comunicativas dessa dimensão são utilizadas para mostrar ao locutor que o ouvinte está acompanhando ou não o que está sendo dito. Frequentemente isso pode ser expressado de maneira não verbal.
Turn Management	engloba funções comunicativas referentes a ações que o locutor toma para assumir alguma posição no diálogo, como manter sua fala, obter fala, desistir de falar e etc.
Time Management	as funções comunicativas dessa dimensão são empregadas quando o locutor pára sua fala por algum momento antes de concluir, seja para ganhar tempo ou por alguma outra razão.
Discourse Structuring	engloba funções comunicativas utilizadas para estruturar um diálogo, como iniciar uma conversa, mudar de assunto, finalizar uma conversa, entre outros.
Own and Partner Communication Management	possui funções comunicativas utilizadas principalmente quando um locutor está corrigindo sua própria fala ou complementando a fala de outra pessoa
Social Obligations Management	as funções comunicativas dessa dimensão são bastante utilizadas em diversos tipos de diálogos diferentes, envolvendo funções como “thanking”, “greeting”, “apology”, entre outros
Contact Management	engloba funções comunicativas utilizadas para checar se a pessoa com quem o locutor deseja fazer contato está presente ou disponível.

Assim como em outras taxonomias para anotação de atos de diálogo, o ISO foi desenvolvido para anotação principalmente de diálogos envolvendo duas pessoas, em que uma das pessoas faz o papel de locutor, em determinado momento do diálogo, e a outra de interlocutor. Entretanto, essa não é a estrutura dos textos jornalísticos do corpus do presente trabalho. Ainda assim, é possível utilizar os atos de fala propostos pelo padrão, mas, devido à sua natureza, as dimensões e funções comunicativas inerentes a diálogos, por exemplo, as das dimensões *Auto-Feedback e Allo-Feedback, Turn Management, Time Management, Discourse Structuring, Own and Partner Communication Management e Contact Management* acabaram não sendo aproveitadas, sendo somente a dimensão *Social Obligation* utilizada. Das 56 funções comunicativas definidas pelo ISO, 17 foram encontradas durante a anotação. As Tabelas 3 e 4 apresentam todas as funções comunicativas de Propósito Geral e Propósito Específico encontradas durante a anotação, respectivamente. Enquanto as Tabelas 5 e 6 apresentam as que não foram utilizadas.

Tabela 3. Funções comunicativas de propósito Geral encontradas durante a anotação

Information providing functions	
Função	Descrição
<i>inform</i>	a sentença tem como objetivo passar uma informação. “Tite diz querer seguir em a seleção após o Mundial de a Rússia.”
<i>agreement</i>	locutor demonstra concordar com algo “Embora essa anistia seja difícil de defender, pela desvantagem imposta a quem cumpriu as regras, foi o compromisso possível na ocasião.”
<i>disagreeent</i>	locutor demostra discordar de algo, opinião contrária “Esse negócio de melhor sorvete de o mundo é bobagem.”
<i>confirm</i>	locutor confirma algo ou alguma informação “Isso mesmo, 4 vezes mais rentável após a privatização.”
<i>disconfirm</i>	locutor nega algo ou alguma informação “Não é verdade o que se fala, que eu não tinha o direito de sair de o país.”
<i>correction</i>	locutor corrige alguma informação “O chope exclusivo é, na verdade, produzido pela cervejaria Quinta do Malte.”
<i>answer</i>	locutor está respondendo algum questionamento A: “E o papel do médico ?”B:“O papel de o médico é de tutela, de referência.”
Information seeking functions	
<i>question</i>	locutor tenta obter uma informação através de um questionamento “Para onde esse ícone vai em o futuro?”
Comissive functions	
<i>promise</i>	locutor se compromete a realizar alguma ação ou que algo acontecerá “Prometo que aprenderei com isso e jamais tomarei qualquer atitude assim mais uma vez.”
Directive functions	
<i>request</i>	locutor solicita algo a alguém “Pessoal, um passinho em o corredor.”
<i>instruct</i>	locutor dá instrução sobre algo

	“Para participar, é preciso acessar o site esaf.fazenda.gov.br , da Escola de Administração Fazendária, ou a página do Tesouro Direto (tesouro.fazenda.gov.br/tesouro-direto).”
<i>suggest</i>	locutor sugere algo “Lave o carro na sombra, para que as substâncias químicas não causem manchas.”

Tabela 4. Funções comunicativas de Propósito Específico encontradas durante a anotação

Social Obligations Managements	
<i>apology</i>	locutor tem a intenção de demonstrar arrependimento “Espero, do fundo do coração, que este ato errado de minha parte seja perdoado.”
<i>thanking</i>	locutor demonstra gratidão por algo “É um orgulho e uma honra receber essa faixa de capitão de a seleção.”
<i>compliment</i>	locutor faz um elogio a algo ou alguém “Obama é uma figura incrível.”
<i>congratulations</i>	locutor parabeniza algo “Parabenizo a Folha pela iniciativa de abrir espaço para manifestação de crianças.”
<i>sympathy expression</i>	locutor demonstra empatia por algo “Tenho acompanhado, com tristeza, o noticiário sobre a lamentosa morte do professor Cancellier.”

Tabela 5. Funções comunicativas de propósito Geral não encontradas durante a anotação

Information seeking functions	
Função	Descrição
<i>propositional question</i>	locutor quer checar se algo é verdade / pergunta de sim ou não “O festival será realizado aqui?”
<i>set question</i>	locutor deseja obter informação sobre algum atributo relacionado à pergunta “Qual o objetivo dos eleitores ao tomarem essa decisão?”
<i>check question</i>	o locutor acredita que uma informação é verdadeira e tenta checar se realmente é “Mas você sabia que isso era errado, né?”
<i>choice question</i>	pergunta com opções que o locutor deseja checar se estão corretas “É para ir pra esquerda ou para direita?”
<i>test question</i>	o locutor faz uma pergunta para checar o conhecimento de alguém Professor: “Quantos estados tem o Brasil?”

Comissive functions	
<i>offer</i>	locutor se oferece para realizar alguma ação caso seja do interesse “Vou checar essa informação para você”
<i>address request</i>	locutor demonstra considerar realizar uma ação requisitada A: “Você consideraria fazer isso por mim?”B: “Se você me ajudar com os documentos, sim”
<i>accept request</i>	locutor aceita realizar uma ação requisitada A: “Com licença, você aceita responder algumas perguntas?”B: “Claro”
<i>decline request</i>	locutor rejeita realizar uma ação requisitada A: “Senhor, você precisa esperar mais alguns minutos.”B: “Não!”
<i>address suggest</i>	locutor demonstra considerar realizar uma ação sugerida A: “O que você acha de provar aquele?”B: “Só se você dividir comigo.”
<i>accept suggest</i>	locutor aceita realizar uma ação sugerida A: “Se eu fosse você começaria por aquela parte”B: “Verdade, vou fazer isso!”
<i>decline suggest</i>	locutor rejeita realizar uma ação sugerida A: “Se quiser você pode esperar aqui”B: “Não, obrigado!”
Directive functions	
<i>address offer</i>	locutor demonstra considerar aceitar uma oferta A: “Pode deixar que eu pago para você”B: ”Mas amanhã eu te devolvo”
<i>accept offer</i>	locutor aceita uma oferta A: “Eu faço pra você”B: ”Obrigado!”
<i>decline offer</i>	locutor rejeita uma oferta A: “Eu faço pra você”B: ”Não, obrigado!”

Tabela 6. Funções comunicativas de Propósito Específico não encontradas durante a anotação

Feedback functions	
Função	Descrição
<i>auto positive</i>	locutor demonstra entender o que está sendo dito “Aham!”
<i>auto negative</i>	locutor demonstra não entender o que está sendo dito “Como?”

<i>allo positive</i>	locutor demonstra que o entendimento de interlocutor ao que foi dito está correto “Isso! Correto!”
<i>allo negative</i>	locutor demonstra que o entendimento do interlocutor em relação ao que foi dito está errado “Não, não, não!”
<i>feedback elicitation</i>	locutor checa se o interlocutor entendeu sua fala “Tudo bem até aqui?”
Turn Management	
<i>turn accept</i>	locutor demonstra aceitar assumir o posto de fala após requisitado A: “O que vc acha?” B: “Ah, deixa eu pensar ...”
<i>turn take</i>	locutor toma o posto de fala em um momento em que teve espaço A: ... B: “Então, ...”
<i>turn grab</i>	locutor pega o espaço de fala de outra pessoa “Desculpa interromper, mas eu penso outra coisa.”
<i>turn assign</i>	locutor tem a intenção de passar a fala para outra pessoa “Matheus, gostaria de complementar?”
<i>turn release</i>	locutor tem a intenção de dar a oportunidade de fala para outras pessoas “Acredito que da minha parte seja tudo!”
<i>turn keep</i>	locutor tem a intenção de manter a fala “Além disso, ...”
Time Management	
<i>stalling</i>	locutor tenta ganhar tempo durante sua fala “Hum, deixa eu pensar!”
<i>pause</i>	locutor indica uma pausa na sua fala por um breve momento “Só um instante!”
Discourse Structuring	
<i>interaction structuring</i>	locutor tem a intenção de estruturar o diálogo “Bom, então acredito que agora podemos falar sobre a próxima tarefa.”
<i>opening</i>	locutor tem a intenção de mostrar que está pronto pra começar o diálogo “Pronto, podemos começar!”
<i>topic shift</i>	locutor demonstra querer mudar de assunto “Mais alguma coisa?”
Own and Partner Communication Management	
<i>self error</i>	locutor mostra que cometeu um erro em sua fala “Na verdade vai ser amanhã.”
<i>retraction</i>	locutor corrige um erro cometido na mesma fala “Aqueles acomodados, digo, advogados, não queriam me ajudar.”

<i>self correction</i>	locutor corrige um erro de formulação cometido na mesma fala “Então, nós íamos, nós vamos começar a implantar essa ideia.”
<i>completion</i>	locutor ajuda o interlocutor a concluir sua fala A: “Acho que a reunião vai ser, deixa eu verificar, ...”B: “10 horas”
<i>correct misspeaking</i>	locutor corrige a fala de outra pessoa A: “Pra realizar a receita você vai precisar de 3 ovos, uma colher de açúcar”B: “Duas colheres”
Contact Management	
<i>contact check</i>	locutor checa se outra pessoa está disponível para conversar “Olá?”
<i>contact indication</i>	locutor demonstra que está pronto para conversar “Pronto”
Social Obligations Managements	
<i>init greeting</i>	locutor tem a intenção de iniciar o diálogo com outra pessoa “Bom dia, caro leitor, na matéria de hoje o assunto será surpreendente.”
<i>return greeting</i>	locutor responde a uma iniciativa de diálogo “Oi, tudo bem e você?”
<i>init self introduction</i>	locutor se apresenta “Prazer, eu sou o dono deste estabelecimento.”
<i>return self introduction</i>	locutor se apresenta após o interlocutor se apresentar para ele A: “Prazer, eu sou o dono deste estabelecimento.”B: “O prazer é meu! Eu sou o vizinho da loja”
<i>accept apology</i>	locutor aceita um pedido de desculpas “Imagina, não foi nada!”
<i>accept thanking</i>	locutor responde a uma demonstração de gratidão de outra pessoa “De nada!”
<i>init goodbye</i>	locutor inicia despedida “Até mais!”
<i>return goodbye</i>	locutor se despede após alguém se despedir A: “Até mais!”B: “Até!”

2.2. Especificação do Procedimento e fluxo de anotação

Para cada sentença do corpus devem ser anotados:

1. Tipo do Grupo: Propósito Geral ou Propósito Específico. Apenas um grupo deve ser escolhido.
2. Dimensão / Agrupamento: Para o grupo de Propósito Geral deve ser identificado o agrupamento da função, sendo eles: *Information Seeking, Information Providing, Comissive ou Directive*. Para o grupo de Propósito Específico deve ser identificada a dimensão da função comunicativa, sendo elas: *Feedback, Turn Management, Time Management, Discourse Structuring, Own and Partner Communication Managment, Social Obligation Management e Contact Management*. Apesar de apenas as funções comunicativas da dimensão *Social Obligation Management* terem sido encontradas durante a anotação, todas elas devem ser consideradas ao realizar a anotação, visto que, apesar de menos provável, todas podem aparecer. Apenas uma dimensão ou agrupamento deve ser escolhido.
3. Função Comunicativa: *inform, agreement, disagreement, confirm, disconfirm, correction, answer, question, propositional question, set question, check question, choice question, test question, promise, offer, address request, accept request, decline request, address suggest, accept suggest, decline suggest, instruct, request, suggest, address offer, accept offer, decline offer, auto positive, auto negative, allo positive, allo negative, feedback elicitation, turn accept, turn take, turn grab, turn assign, turn release, turn keep, stalling, pausing, topic shift, interaction structuring, opening, retraction, self error, self correction, completion, correct misspeaking, contact check, contact indication, init greeting, apology, thanking, compliment, congratulation, sympathy expression, return greeting, init self introduction, accept apology, accept thanking, init goodbye, return goodbye*. Apenas uma função comunicativa deve ser escolhida.

Para isso, os seguintes passos devem ser realizados:

1. Entendimento do corpus: o primeiro passo de preparação para a anotação do corpus é realizar uma breve análise de suas sentenças. Para isso, o anotador deve selecionar aleatoriamente um pequeno conjunto de notícias, cerca de 10 pelo menos, e realizar a leitura de todas as sentenças para criar familiaridade como o corpus.
2. Estudo das dimensões e funções comunicativas do ISO: após isso, o anotador deve realizar o estudo das dimensões e funções comunicativas definidas pelo ISO-24617-2. Para isso é necessário realizar a leitura do documento [ISO 2012], focando principalmente nas partes em que são explicadas cada uma das dimensões e funções e que contêm instruções sobre a anotação. Espera-se que durante esse estudo o anotador exercite assimilar as explicações das funções com as sentenças que analisou do corpus. Isso ajudará a desenvolver uma intuição sobre as dimensões e funções comunicativas que serão mais utilizadas e também as dimensões que dificilmente aparecerão no corpus. É recomendada a leitura da explicação de todas as dimensões e a análise de cada uma das funções comunicativas que elas englobam, assim como os exemplos das unidades textuais correspondentes a cada função. Durante essa etapa, o anotador irá notar que, em muitos casos, as especificações das funções comunicativas possuem uma explicação voltada principalmente para diálogos, em que existe um locutor e um interlocutor. No processo

de anotação, o anotador humano deve se colocar na função de interlocutor, uma vez que o escritor/produtor das notícias que compõem o corpus jornalístico em questão tem como finalidade descrever ou apresentar um fato ou evento ao leitor, tendo, nesse cenário, a função de locutor.

Passos de anotação:

1. Identificar a dimensão: Para tanto o anotador deve considerar:
 - (a) se a sentença faz uma narrativa, passa alguma informação ou procura obter alguma informação;
 - (b) o entendimento do locutor ou do ouvinte das sentenças anteriores;
 - (c) a atribuição do papel de orador ou locutor (como a tentativa de ocupar ou passar o espaço de fala);
 - (d) o tempo necessário para continuar o diálogo (como existência de pausas durante falas e a intenção dessas pausas);
 - (e) as correções do que o interlocutor ou locutor está dizendo (como uma auto correção de algo mencionado ou a correção da fala de algum terceiro);
 - (f) a estrutura do diálogo (como início e fim de conversas e trocas de assuntos);
 - (g) a tentativa de contato (como verificar se o interlocutor está disponível para um diálogo);
 - (h) obrigações sociais (como agradecimentos).

Algumas dimensões mais específicas de diálogos podem não aparecer ou serem menos recorrentes. Entretanto, em algumas exceções, como menções de trechos de conversas, essas dimensões podem ser utilizadas. Para a identificação das dimensões e agrupamentos, deve-se seguir as descrições e considerações contidas nas Tabelas 1 e 2. Após a identificação da dimensão, deve-se escolher a função comunicativa da sentença.

2. Identificar a função comunicativa: para a identificação das funções de propósito geral, o anotador deve se basear nas descrições e exemplos contidos nas Tabelas 3 e 5 e, para identificar as funções de propósito específico, o anotador conta com as Tabelas 4 e 6. Vale ressaltar que, devido ao corpus não se tratar de um diálogo, a forma que cada função comunicativa foi interpretada e exemplificada nas tabelas pode destoar da maneira explicada pelo ISO, uma vez que suas interpretações são mais voltadas para diálogos.

Além disso, durante a anotação, deve-se considerar não apenas a sentença que está sendo anotada, mas também o contexto em que ela está sendo empregada. Como no corpus há sentenças faltando, ao mudar a anotação entre uma sentença e outra de uma mesma notícia, é preciso observar o ID da sentença para saber se são sentenças consecutivas e evitar desta forma tomar conclusões precipitadas à respeito do seu contexto.

Uma vez que a dimensão de propósito geral tenha sido identificada, é possível se basear na árvore da Figura 2 para escolher a função comunicativa mais adequada que, no caso, deve ser a mais específica possível. Isso pode ser feito percorrendo-se a árvore no sentido *top-down*. Em um primeiro momento, podemos olhar para a sentença “*Isso é uma vergonha para os nova-iorquinos.*”, por exemplo, e perceber que se trata da função comunicativa “inform”. Porém, ao mesmo tempo, representa uma discordância. O ideal é buscar escolher sempre a função comunicativa mais específica para a sentença (neste caso, seria a etiqueta “disagreement”).

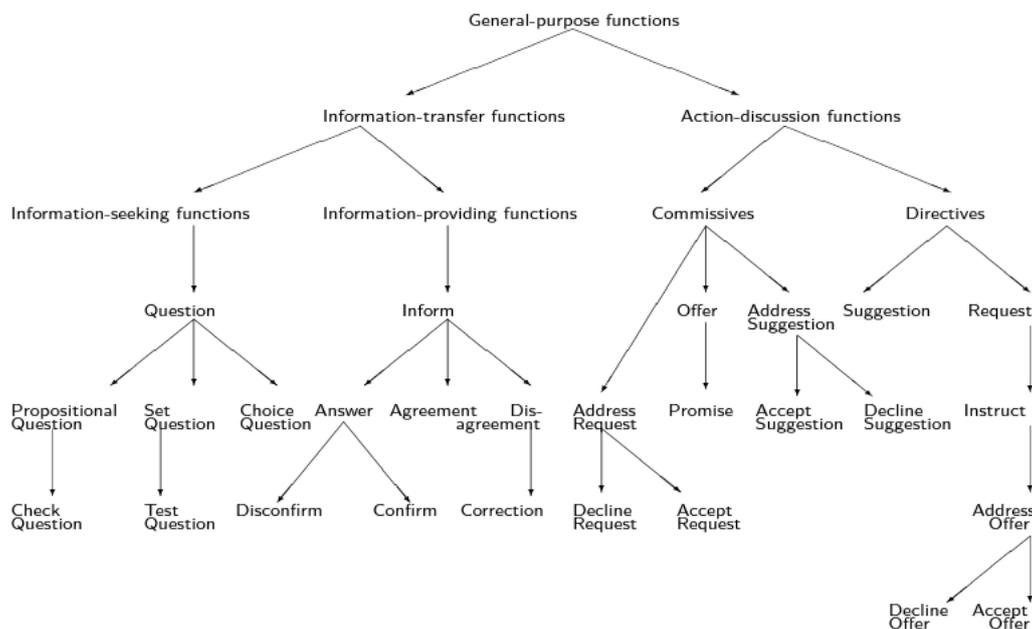


Figura 2. Funções de propósito Geral
ISO/DIS 24617-2

Além disso, é possível que ocorram casos em que mais de uma etiqueta seja adequada para a anotação. Na sentença “*Até resolver esse problema filosófico, convém continuar a investir em métodos anticoncepcionais.*”, por exemplo, é possível identificar duas funções comunicativas: “inform” e “suggestion”. Nessa situação, fica a cargo do anotador a seleção do rótulo a ser utilizado para anotar a sentença em questão.

Outro ponto de atenção diz respeito à anotação de sentenças que informam ou narram um ato de fala. A sentença “*Ao saber que teria que abandonar a prova, Vettel pediu desculpas à equipe.*”, por exemplo, narra um pedido de desculpas, porém a mesma não possui a função comunicativa de um pedido de desculpas, mas sim de informar sobre esse ato.

Durante a anotação de cada uma das sentenças o anotador também poderá sinalizar se teve dúvida. Ao final, o anotador deve retornar para esses casos de dúvida para atribuir uma anotação final.

2.3. Tempo estimado para anotação

Espera-se que de início a anotação leve um pouco mais de tempo, enquanto o anotador se acostuma com a forma das sentenças do corpus e também com a taxonomia utilizada para a anotação. Após esse primeiro momento, é esperado que o anotador crie maior familiaridade com o processo de anotação e consiga anotar cerca de 500 sentenças por hora, sem interrupção.

2.4. Demonstração da anotação

A anotação foi realizada utilizando uma planilha online compartilhada entre os membros do projeto. Cada linha da planilha correspondendo à uma sentença do corpus com

sua identificação e a identificação da notícia em que está inserida. A Tabela 7 apresenta um exemplo da forma como a anotação foi realizada.

Tabela 7. Ilustração da anotação no corpus

ID notícia	ID sentença	sentença	Tipo	Conjunto / Dimensão	Função
1	3	‘Eu sei que tô lascado , todo dia tem um processo .’	Geral	information providing	disagreement
1	4	Eu não quero nem que Moro me absolva , eu só quero que ele peça desculpas , disse Lula durante um seminário sobre educação em Brasília .	Geral	information providing	inform

Agradecimentos

Este projeto foi apoiado pelo Ministério da Ciência, Tecnologia e Inovações, com recursos da Lei nº 8.248, de 23 de outubro de 1991, no âmbito do PPI-SOFTEX, coordenado pela Softex e publicado Residência em TIC 13, DOU 01245.010222/2022-44.

Os autores deste trabalho agradecem ao Centro de Inteligência Artificial (C4AI-USP) e o apoio da Fundação de Apoio à Pesquisa do Estado de São Paulo (processo FAPESP #2019/07665-4) e da IBM Corporation.

Referências

- Austin, J. L. (1962). *How to do things with words*. Oxford University Press, 1 edição.
- Folha, G. (1921). Folha de são paulo. [urlhttp://www.folha.uol.com.br/](http://www.folha.uol.com.br/) .
- ISO (2012). Iso 24617-2:2012: Language resource management – semantic annotation framework (semaf) – part 2: Dialogue acts.
- Lopes, L., Duran, M. S., Nunes, M. d. G. V., e Pardo, T. A. S. (2022). Corpora building process according to the universal dependencies model: an experiment for portuguese.
- Nivre, J., de Marneffe, M.-C., Ginter, F., Hajič, J., Manning, C. D., Pyysalo, S., Schuster, S., Tyers, F., e Zeman, D. (2020). Universal Dependencies v2: An evergrowing multilingual treebank collection. Em *Proceedings of the Twelfth Language Resources and Evaluation Conference*, páginas 4034–4043, Marseille, France. European Language Resources Association.
- Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, London.