

Mensurando Saliência: uma medida com base em ênfases na agenda legislativa do Brasil (1995-2014)

Mariana Batista

Bhreno Vieira

Universidade Federal de Pernambuco

Universidade Federal de Pernambuco

Primeira Versão, 2016

Resumo

Quais são os temas mais enfatizados na agenda legislativa brasileira? Diferentes atores políticos enfatizam diferentes áreas temáticas? Esse trabalho explora as ênfases temáticas dos atores políticos com base na classificação das proposições legislativas na Câmara dos Deputados entre 1995 e 2014. O objetivo é apresentar uma classificação específica e compreensiva para identificar as diferenças de ênfases e estabelecer a saliência temática para o presidente e para os diferentes partidos políticos. Para tanto, o trabalho utiliza a tipologia proposta pelo “The Comparative Agendas Project”. A metodologia usa análise de conteúdo implementada com “aprendizado supervisionado” para classificar as iniciativas legislativas apresentadas na Câmara dos Deputados. Uma vez realizada a classificação, a medida de saliência é construída através do número de iniciativas propostas por cada ator em determinado tema. A proposta contribui para o debate ao oferecer uma medida de saliência para os principais atores políticos ao longo do tempo, podendo ser usada em trabalhos desde relação executivo-legislativo a embate de interesses no Congresso Nacional.

Palavras-Chave: Saliência, Ênfases, Agenda Legislativa, Análise de Conteúdo, Machine Learning.

Artigo preparado para o 10 Encontro da Associação Brasileira de Ciência Política, 30 de Agosto a 2 de Setembro, 2016, Belo Horizonte, MG.

1 Introdução

Na definição clássica de Laswell política é quem ganha o quê, quando e como (Lasswell, 1950). A partir dessa definição, muitos trabalhos foram realizados para entender os seus elementos: quem são os atores envolvidos? O que diferencia eles? Quais são os recursos mobilizados? E mais recentemente, quais são as regras desse jogo? No presente trabalho focamos no conteúdo dessa barganha política. Isto é, sobre o que exatamente essa disputa se desenvolve.

Atores políticos buscam influência sobre decisões, e essas decisões se referem a determinados temas ou áreas temáticas. Disputas políticas se dão em torno dos mais diferentes temas: políticas de saúde, de educação, segurança pública, direitos de minorias, regulação da indústria, dentre um sem número de outros. Tendo isto em mente, é possível analisar esses temas como sendo todos iguais? Corriqueiramente classificamos os temas de acordo com a sua importância, sua visibilidade ou grau de prioridade. Dizemos que algumas decisões são mais importantes do que outras, ou que determinadas políticas são mais visíveis ou são prioridade. Além das diferenças entre os temas, também podemos observar diferenças de percepção dos atores. Por exemplo, esperamos que políticas que se referem ao meio ambiente sejam mais importantes para membros de um partido verde, ou que partidos de direita defendam uma agenda conservadora ou ainda que parlamentares ligados a um grupo de interesse específico se preocupe especialmente com o tema dessa grupo.

Usando diferentes termos, o que estamos dizendo é que os diferentes temas tem diferença na sua saliência ou na sua importância relativa (Humphreys & Garry, 2000). Alguns temas são mais salientes do que outros e alguns temas são mais salientes para um determinado ator político do que para outro. O objetivo desse trabalho não é explicar essas diferenças, mas apresentar uma classificação temática que ofereça elementos para que esses outros argumentos possam ser testados.

Saliência é algo que existe um relativo consenso em torno de sua relevância, mas grandes dificuldades na sua mensuração e observação empírica. Medidas recentes de ênfases temáticas vem sendo construídas com base em surveys (Laver & Hunt, 1992, Warwick & Druckman, 2006), manifestos partidários (Laver *et al.* , 2003b), discursos políticos (Laver *et al.* , 2003b, Quinn *et al.* , 2010) e comunicações públicas (Grimmer, 2013). Propomos aqui uma medida de saliência com base nas ênfases temáticas da agenda legislativa da Câmara dos Deputados no período de 1995 a 2014. A tipologia utilizada para a classificação dos temas é a proposta pelo "Comparative Agendas Project" (Baumgartner *et al.* , 2011), que apresenta 20 tópicos e mais

de 200 subtópicos, desde temas gerais como educação e saúde a temas bem mais específicos como direitos das mulheres e regulação de planos de saúde. Propomos como medida de saliência a frequência de iniciativas legislativas propostas por cada ator político em cada um dos temas da tipologia.

Utilizar dados da agenda legislativa se mostra apropriado porque oferece uma medida que pode ser analisada no tempo e também por ator específico, desde o presidente, passando pelos partidos políticos, até o parlamentar individual. Como a agenda legislativa da Câmara dos Deputados no período de 1995 a 2014 conta com mais de 40 mil proposições legislativas, implementamos uma estratégia baseada na codificação manual (hand-coding) de uma amostra das proposições e o uso dessa amostra para a classificação do total de iniciativas a partir de uma técnica de "aprendizado supervisionado" (supervised learning), uma aplicação de machine learning. Essa estratégia se mostra mais eficiente em termos de tempo de classificação e reduz a subjetividade no processo classificatório, ao mesmo tempo que mantém a qualidade do resultado final.

Acreditamos que o artigo oferece uma classificação que pode ser utilizada para responder questões de grande importância para a Ciência Política brasileira como a atuação de grupos de interesse no Legislativo, relações Executivo-Legislativo e também sobre a conexão entre demandas da sociedade e atuação legislativa. Adicionalmente, implementa uma metodologia que acelera o processo de classificação de grandes quantidades de textos desestruturados, mantendo a precisão.

O restante desse artigo está assim organizado: a próxima seção apresenta a definição de saliência e as possibilidades de mensuração. Em seguida discutimos a coleta de dados, a tipologia utilizada, o processo de classificação da amostra e o algoritmo de classificação do universo de proposições legislativas da Câmara dos Deputados no período de 1995 a 2014. Depois apresentamos os resultados preliminares de aplicação da técnica e por último a conclusão.

2 Caminhos para o Estudo de Ênfases Temáticas

A competição política é um dos temas de maior atenção na Ciência Política. Contudo, "less attention is given to to what it is that parties are competing about" (Laver & Hunt, 1992, 1). Partidos políticos disputam o voto do eleitor em termos de políticas ou de temas específicos. Apesar de frequentemente nos referimos a essas políticas de forma bastante generalista e de análises de competição política constantemente analisar o conteúdo do que se está decidindo

de maneira bastante superficial, "it quickly becomes clear when we begin to develop more precise models of political competition that the 'policy' of any given actor has a number of quite distinct elements" (Laver, 2001, 66).

Esses elementos que distinguem as políticas normalmente se relacionam a sua diferença em importância, relevância ou prioridade. Todos esses diferentes termos se referem à noção comum de que alguns temas são mais importantes do que outros e essa importância varia no tempo e também entre atores. Em suma, essas diferenças remetem à noção de saliência. Saliência pode ser definida como "the relative importance of different policy areas" (Humphreys & Garry, 2000, 2). *Issue Salience* então se constitui num tema fundamental para as teorias sobre competição política (Robertson, 1976).

Saliência é relacionada à outra característica importante das políticas que é a posição. Saliência se refere a quão importante um determinado tema é para um ator e posição diz se o ator é a favor ou contra. De acordo com Laver (2001), saliência e posição são conceitos distintos dado que dois atores podem se importar intensamente com um determinado tema e um ser a favor e o outro ser contra. Por exemplo, dois parlamentares podem ser igualmente preocupados com a legalização do aborto, mas um ser a favor e o outro contra. De acordo com Humphreys & Garry (2000), essa distinção entre saliência e posição é a base da abordagem clássica da saliência. Uma alternativa a essa abordagem vem da abordagem da valência que identifica que em algumas situações onde há consenso quanto à direção, a importância dada ao tema determina também a posição. Nesses casos, "issue salience is position" (Humphreys & Garry, 2000, 3).

Dessa forma, a competição política gira em torno de políticas sobre quais os atores tem preferências. A estrutura dessas preferências tem a ver com a posição ou ponto ideal e também com a importância dada a cada política num contexto multidimensional. Em suma, a saliência é a característica de importância ou intensidade dessas preferências e a posição se refere à direção das mesmas. A posição, é extremamente explorada na literatura, principalmente com o desenvolvimento de modelos espaciais e medidas de ideologia. Já a saliência teve menor desenvolvimento em comparação. Por essa razão, o foco do presente trabalho está na saliência das diferentes políticas e não na sua posição. Isso quer dizer que nos preocupamos aqui com a importância ou ênfase dada a determinados temas pelos atores políticos, mas que não podemos dizer nada sobre como esses atores se posicionam com relação a esses temas (Laver, 2001).

Como a saliência vem sendo observada em trabalhos empíricos? Apesar da centralidade desse conceito para as teorias formais de competição política, não há um consenso com relação

à como observá-la. Esse déficit gerou o problema identificado por Laver & Hunt (1992) de que as teorias se tornaram cada vez mais complexas e essa complexidade não foi acompanhada pelo desenvolvimento de formas de mensuração apropriadas. Passados mais de vinte anos desse diagnóstico, diferentes estratégias foram propostas para lidar com a complexidade do problema. Estas estratégias envolvem a utilização de surveys, a codificação de manifestos partidários, a análise de discursos políticos ou de outras comunicações públicas. O ponto em comum é saber quais são os temas enfatizados ou mais importantes para cada ator político e nos diferentes períodos de tempo.

Uma primeira estratégia para a mensuração da saliência das políticas para os atores políticos é com o uso de surveys. Esses surveys podem ser surveys em massa para identificar a ênfase em determinados temas a partir de perguntas do tipo "qual o problema mais importante". Esse tipo de estratégia vem sendo extremamente criticada por confundir a importância do tema com o seu reconhecimento como um "problema" (Wlezien, 2005). Especificamente com relação a atores políticos, uma segunda estratégia seria realizar survey diretamente com essas elites. Apesar dessa ser a melhor opção em termos de survey uma vez que identifica a saliência dos temas a partir da percepção do próprio ator, esses surveys são normalmente limitados devido a dificuldade de conseguir uma taxa de resposta que permita estimar a saliência dos partidos políticos, principalmente os partidos menores.

A principal estratégia para a implementação de surveys para identificação da saliência relativa dos temas tem sido o survey com especialistas. A principal contribuição nesse sentido foi o survey conduzido com cientistas políticos por Laver & Hunt (1992). Nesse survey os autores perguntaram a cientistas políticos em 25 sistemas políticos diferentes sobre o posicionamento e a importância de cada tema para cada partido político. Especificamente sobre a questão da saliência, os respondentes foram pedidos para atribuir um escore de "importância" sobre cada dimensão para cada partido no sistema político. Esse survey foi a base para diversos trabalhos sobre governos de coalizão em sistemas parlamentaristas. Ainda com base nesse survey foram desenvolvidas medidas de importância dos diferentes ministérios, relacionando-os às áreas temáticas (Laver & Shepsle, 1994, Warwick & Druckman, 2001, Warwick & Druckman, 2006).

A utilização de surveys com especialistas em muitas situações pode ser a solução mais apropriada. Contudo, essa solução tem uma grande limitação que é o tempo. Esses surveys levam tempo para ser concluídos e dificilmente podem ser aplicados anualmente. Dessa forma, estimativas realizadas num período de tempo são utilizadas muitos anos depois, devido à impossibilidade de atualização contínua das estimativas.

Uma estratégia alternativa é a análise de conteúdo de documentos políticos. Como os atores políticos podem ser responsabilizados pelas promessas que fazem, os documentos que expressam a suas ênfases temáticas ou prioridades se tornam matéria-prima principal para a análise da saliência para esse determinado ator. Uma versão dessa abordagem extremamente bem-sucedida é o Manifesto Project¹. Através da análise de conteúdo das ênfases programáticas dos manifestos partidários é possível identificar quais temas são mais relevantes ou mais enfatizados por cada partido político (Budge, 2001a,b, Laver *et al.*, 2003a). Essa abordagem é especialmente interessante porque se baseia no auto-posicionamento pelos próprios atores políticos e não numa classificação por terceiros. Atualmente o projeto já classificou mais de 1.000 partidos em mais de 50 países desde 1945.

O esquema de classificação padrão envolve 7 domínios e 54 categorias. Os domínios são: Relações Exteriores, Liberdade e Democracia, Sistema Político, Economia, Estado de Bem-Estar Social e Qualidade de Vida, Sociedade e Grupos Sociais (Volkens, 2001). A identificação das ênfases é feita com base na proporção dos textos dedicada aos diferentes temas. A classificação do Manifesto Project é usada não só para a identificação dos temas mais enfatizados, mas também para a localização da posição dos partidos políticos no espectro esquerda-direita a partir da ênfase em determinados temas. O processo de codificação é extremamente regulado, com a recomendação clara de mais de um codificador por documento a ser codificado. Grande parte do trabalho do Manifesto Project foi feito com base em codificação manual e com o cálculo de medidas de confiabilidade para reduzir a subjetividade na classificação.

Recentemente foi desenvolvido um pacote para automatizar a classificação desses documentos e assim tornar a atividade mais eficiente². No Brasil essa abordagem vem sendo desenvolvida nos trabalhos de Tarouco (Tarouco, 2011, Tarouco & Madeira, 2013), que apresenta a aplicação do esquema de codificação tradicional para os partidos brasileiros e também um esquema de classificação adaptado para as especificidades do país.

Uma limitação do uso de manifestos partidários como medida de saliência pode ser relacionada à heterogeneidade desses documentos, que muitas vezes não detalham os temas de interesse. Outro ponto é a limitação temporal da disponibilidade dessa informação. Isto é, manifestos partidários são apresentados no momento de fundação do partido e essas ênfases são diretamente relacionadas a esse momento, sendo limitadas para descrever as prioridades atuais. Essa limitação pode ser minimizada ao utilizar manifestos eleitorais como documento

¹<https://manifestoproject.wzb.eu/>

²<https://manifestoproject.wzb.eu/information/documents/manifestoR>

principal, mas esses manifestos também se referem ao momento específico eleitoral e nem todos os candidatos divulgam esse documento.

Uma outra opção na identificação de ênfases ou da saliência de temas que resolve a limitação das medidas anteriores quanto à disponibilidade de informações anualmente e também da relação custo-benefício é o foco nos pronunciamentos públicos. Esses pronunciamentos podem ser discursos em plenário, comunicações à eleitores ou ainda comunicados à imprensa. Por muito tempo essa opção não foi explorada a contento devido às limitações tecnológicas para sua análise. Isso porque o material a ser analisado desses pronunciamentos acaba sendo milhares de páginas de transcrições, para as quais a codificação humana se mostrava muito custosa. Com o desenvolvimento da análise automatizada de conteúdo essa opção se mostrou mais viável e diversos trabalhos vem sendo desenvolvidos para identificar as ênfases temáticas dos políticos. Quinn *et al.* (2010) utiliza um método automatizado baseado em dicionários para identificar a atenção a diferentes tópicos a partir dos discursos dos parlamentares americanos. Outra contribuição nesse sentido está no trabalho de Grimmer (2013) que analisa as comunicações políticas dos senadores americanos para identificar a "agenda expressa" desses parlamentares. No Brasil, essa abordagem inovadora foi desenvolvida por Moreira (2016) que aplicou o modelo de Grimmer (2013) para identificar as ênfases temáticas nos discursos dos deputados brasileiros. O trabalho analisa mais de 127 mil pronunciamentos num período de mais de 15 anos.

O presente trabalho segue a linha de identificar a saliência temática a partir da atuação dos próprios atores políticos. Mas ao invés de focar no que os atores políticos dizem, resolvemos enfatizar aquilo que fazem quando tem liberdade para fazê-lo. Analisamos a importância dada aos diferentes temas de políticas a partir da agenda de políticas dos atores. Ou mais precisamente, a partir daquilo que os atores políticos propõem no processo legislativo.

A agenda de políticas públicas normalmente não é utilizada como medida de preferência dos atores políticos porque é identificada como o produto final de um processo de barganha entre os atores no processo de formação da agenda (agenda-setting). Num dos trabalhos pioneiros sobre agenda de políticas públicas, Kingdon (2003) discute o complexo processo de formação da agenda de políticas, identificando três fluxos que em conjunto aumentariam a possibilidade de um tema ser considerado seriamente pelos atores governamentais. Desenvolvendo o chamado modelo de equilíbrio pontuado, Baumgartner & Jones (2010) discutem a formação da agenda e porque longos períodos de estabilidade são interrompidos por mudanças bruscas. Nota-se a partir dessas duas abordagens que a agenda de políticas é vista como um processo complexo no qual dos temas que interessam aos atores específicos, apenas alguns podem ser considerados

seriamente e somente esses comporiam a agenda de políticas.

O que propomos aqui não é a análise do que é aprovado em termos de políticas no processo legislativo, nem das propostas apresentadas quais conseguem avançar no processo de agenda-setting. Concordamos que esses dois pontos são resultado não só das preferências dos atores, mas também dos processos de trocas entre os mesmos e também do desenho institucional que estabelece que alguns atores tem mais poder para fazer as propostas avançarem do que outros. O foco deste trabalho está no que os políticos propõem como medida para a importância que dão a um determinado tema. Essa abordagem é viável devido às regras de organização do processo legislativo no Brasil.

De acordo com Figueiredo & Limongi (2001), o processo de proposição no Legislativo é descentralizado. Isso porque o parlamentar individual tem poder quase ilimitado de proposição. Contudo, a capacidade de influenciar o resultado legislativo é muito pequena. Isso faz com que a proposição legislativa possa ser observada como um atividade de sinalização de preferências, já que não há maiores constrangimentos para o que pode ser proposto. Para além dos parlamentares, no Brasil o presidente também tem poder de iniciar proposições legislativas, de forma que teríamos uma unidade comum para comparar tanto os parlamentares quanto o presidente. Tendo isto em mente, analisaremos as proposições legislativas como forma de observar a importância ou saliência temática para os diferentes atores políticos no Brasil. Observar saliência através da proposição legislativa tem quatro benefícios principais: 1) são dados publicamente disponíveis; 2) proposições legislativas se referem à políticas e são instrumentos importantes para os atores políticos; 3) não há maiores constrangimentos quanto ao conteúdo do que é proposto; 4) podemos construir medidas de saliência para o indivíduo ou agregados partidários.

Para tanto, utilizaremos o esquema de classificação inicialmente proposto por Baumgartner & Jones (2010) e depois desenvolvido no Comparative Agendas Project³. De forma similar ao Manifestos Project, o Comparative Agendas Project oferece uma tipologia compreensiva que pode ser implementada em diferentes contextos. Utilizamos esse esquema classificatório para identificar quais são os temas mais enfatizados na agenda legislativa brasileira, a partir do poder de proposição de políticas. A próxima seção apresenta a metodologia e em seguida os resultados preliminares da aplicação dessa abordagem.

³<http://www.comparativeagendas.net/>

3 Mensurando Saliência: dados e métodos

Propomos mensurar a saliência dos diferentes temas no Brasil a partir da proposição legislativa dos atores políticos no período de 1995 a 2014. Proposições legislativas se constituem em textos e esses são o material básico para a análise. O foco está nos temas que são enfatizados pelos parlamentares individualmente, pelos partidos políticos como conjuntos de parlamentares e também pelos presidentes. O objetivo é a classificação de cada proposição de acordo com o seu tema predominante. Portanto, uma técnica de análise de conteúdo deve ser implementada.

A análise de conteúdo pode ser definida como "a technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use" (Krippendorff, 2004, 18) ou "as the systematic, objective, quantitative analysis of message characteristics" (Neuendorf, 2002, 1). O foco está em como transformar textos desestruturados em informação que pode ser utilizada para descrever uma realidade. A análise de conteúdo é um termo comum e como técnica pode ser implementada de diferentes formas e com diferentes objetivos. Contudo, como técnica a análise de conteúdo tem procedimentos que devem ser seguidos para garantir o seu rigor.

Neuendorf (2002) apresenta os passos que devem ser seguidos para a implementação da análise de conteúdo: 1) Teoria; 2) Conceitualização; 3) Operacionalização; 4) Técnica de Codificação; 5) Amostragem; 6) Treinamento; 7) Codificação; 8) Cálculo de confiabilidade; 9) Apresentação dos resultados. Apresentaremos a seguir cada uma das escolhas feitas no presente trabalho quanto a cada um desses passos.

Discutimos na seção anterior a teoria e a conceitualização que orienta a coleta de dados. O foco está na saliência como conceito a ser observado através da classificação em temas (issue areas) das proposições legislativas, que se constituem na nossa unidade de análise. Como a unidade de análise é a proposição, o texto a ser codificado é o da ementa da proposição. A operacionalização tem como principal preocupação a construção de uma tipologia que seja simultaneamente exaustiva e suas categorias mutuamente exclusivas. Exaustivo aqui quer dizer que nenhum tema importante foi ignorado. Mutuamente exclusivo quer dizer que nenhum tema pode ser classificado em mais de uma categoria. Para lidar com o problema de construir uma tipologia que seja exaustiva e mutuamente exclusiva, além de garantir a transparência e a replicabilidade dos resultados e em última instância permitir a comparabilidade com resultados de outros materiais e outros contextos utilizamos a tipologia disponibilizada pelo Comparative

Agendas Project⁴.

Decidimos utilizar a categorização do Comparative Agendas Project porque oferece uma classificação compreensiva e porque enfatiza temas importantes para a discussão sobre ênfases temáticas no Brasil. O uso dessa classificação também facilita o trabalho uma vez que se trata de uma categorização já implementada em diversos contextos diferentes. Inicialmente criada por Baumgartner & Jones (2010) para analisar a composição de agenda de políticas nos Estados Unidos, a classificação foi adaptada para uma abordagem comparada, sendo constituído o Comparative Agendas Project. Nesse contexto foram criados projetos locais com o objetivo de implementar a classificação, tendo como objeto não só agendas legislativas, mas também pronunciamentos e dados da mídia. Atualmente 22 sistemas políticos mantêm parceria com o Comparative Agendas Project, inclusive o Brasil⁵.

Quando apresentou a classificação inicial, Baumgartner & Jones (2010) propuseram que as categorias e a teoria andariam juntas. No caso a teoria seria o modelo do equilíbrio pontuado. Não nos filiamos diretamente aos pressupostos teóricos desse modelo, até porque o foco do trabalho não é processo de formação da agenda. Assim como Green-Pedersen & Walgrave (2014), acreditamos que o esquema de classificação não tem uma teoria. Apenas utilizamos a classificação proposta como forma de observar as ênfases temáticas dos atores políticos brasileiros por meio de sua atividade de proposição legislativa.

O esquema de classificação original conta com 20 categorias e mais de 200 subcategorias. Nós traduzimos as categorias e suas respectivas definições e também adaptamos algumas das subcategorias para o contexto específico brasileiro. As modificações foram mantidas ao mínimo para manter a comparabilidade do esquema de classificação e a maior parte das mudanças se deram na categoria de política internacional, uma vez que muitas das categorias são específicas dos países. A classificação final foi feita com base no codebook do projeto americano e do projeto espanhol, ambos disponíveis no site do Comparative Agendas Project. A tabela abaixo apresenta as categorias. Por questão de espaço omitimos as subcategorias.

⁴O codebook original com todas as categorias pode ser encontrado aqui: <http://www.comparativeagendas.net/pages/master-codebook>

⁵<http://www.comparativeagendas.net/brazil>

Tabela 1: Esquema de Classificação Temática

Categorias
1. Macroeconomia
2. Direitos Civis Gerais, Liberdades Civis e Direitos de Minorias
3. Saúde
4. Agricultura
5. Trabalho e Emprego
6. Educação
7. Meio Ambiente
8. Energia
9. Imigração
10. Transporte
11. Questões de Lei, Crime e Família
12. Bem-Estar Social
13. Planejamento Urbano e Políticas Habitacionais
14. Bancos, Finanças e Comércio Interno
15. Defesa
16. Ciência, Tecnologia e Comunicação
17. Comércio Exterior
18. Política Internacional e Ajuda Externa
19. Governo e Administração Pública
20. Recursos Naturais e Gestão da Água

Fonte: elaboração dos autores com base no Comparative Agendas Project Codebook.

Dentro dessas 20 categorias o esquema de classificação conta com 237 subcategorias, que envolvem temas como discriminação racial e de gênero, regulamentação da indústria de remédios, subsídios governamentais para agricultores e benefícios trabalhistas. A categoria geral serve como indicativo da importância daquele tema na agenda legislativa. Já as categorias específicas podem ajudar a responder questões importantes ainda em aberto, como a expansão dos direitos de minorias ou a atuação de grupos de interesse. Acreditamos que essa classificação seja exaustiva e o processo de classificação torna as categorias mutuamente exclusivas, já que uma proposição legislativa só pode ser classificada em uma das subcategorias.

Definida a operacionalização passamos agora à discussão da técnica de categorização. A codificação pode ser manual (hand-coding) ou automatizada com o uso de algum software especializado. Cada opção tem os seus prós e contras. A codificação manual tem como ponto positivo a precisão na classificação. Quando lemos uma proposição legislativa temos a maior capacidade de discernir sobre o que a proposição trata. O mesmo não acontece com a codificação automatizada, que tem que se basear em algum elemento como um dicionário para

identificar à qual tópico a proposição se refere. Contudo, a codificação manual pode ser sujeita à subjetividade, uma vez que a percepção de um indivíduo pode ser diferente de outros. Por essa razão que o chamado padrão de ouro da análise de conteúdo é a codificação por dois codificadores independentes para que depois a concordância entre eles possa ser avaliada. Basear a codificação em múltiplos codificadores reduz o problema da subjetividade, mas só aumenta uma outra limitação da codificação manual: o custo. Isso porque classificar conteúdo de textos leva tempo e considerando que o volume desses textos vem aumentando cada vez mais, o tempo necessário para classificar todo o material pode ser proibitivamente longo. Ademais, a inclusão de mais de um codificador não reduz o tempo necessário para a classificação, já que ambos tem que classificar as mesmas unidades para que a concordância seja calculada.

Por essa razão, os principais desenvolvimentos na análise de conteúdo tem sido com relação à automatização do processo. Utilizar um software especializado para a classificação reduz o custo da atividade já que diminui o tempo necessário e também porque não dependente de codificadores treinados. Contudo, a classificação automatizada perde em precisão uma vez que uma máquina não tem o mesmo discernimento de um ser humano ao ler um texto.

”Rather than replace humans, computers amplify human abilities. The most productive line of inquiry, therefore, is not in identifying how automated methods can obviate the need for researchers to read their texts. Rather, the most productive line of inquiry is to identify the best way to use both humans and automated methods for analyzing texts” (Grimmer & Stewart, 2013, 4).

Tendo em mente que a codificação manual oferece maior precisão, mas que métodos automatizados permitem a codificação de maiores quantidades de texto a um menor custo, a abordagem aqui utilizada combina tanto a codificação manual quanto a codificação automatizada através de um método de aprendizado supervisionado. Aprendizado supervisionado (supervised learning) combina a codificação manual de uma amostra dos dados para ”ensinar” a máquina a codificar o restante do corpus. Essa é uma aplicação específica de aprendizado de máquina (machine learning), que permite o desenvolvimento de algoritmos para que a máquina ”aprenda” ou melhore o seu desempenho em uma determinada tarefa (Grimmer & Stewart, 2013).

A tarefa aqui em questão é a codificação das proposições legislativas. O aprendizado tem por base a utilização de uma amostra das proposições codificadas manualmente (training-set) para que a máquina identifique as palavras associadas a cada uma das categorias e utilize essa

informação para codificar as demais proposições no corpus. Essa abordagem é superior a outras abordagens automatizadas como o uso de dicionários porque permite lidar com classificações com um grande número de categorias, minimizando o erro de classificação. Como o esquema de categorização utilizado aqui possui mais de 200 categorias, utilizar um dicionário de palavras seria muito custoso e muito propenso ao erro. Por essas razões, a classificação das proposições legislativas aqui realizada implementa a abordagem desenvolvida por Purpura & Hillard (2006) e Hillard *et al.* (2008) para os dados da agenda legislativa do Congresso Americano.

Definida a técnica de codificação, passamos para a descrição do corpus. O corpus é a população de textos a ser analisado (Grimmer & Stewart, 2013) e no nosso caso é o total de proposições legislativas apresentadas na Câmara dos Deputados no período de 1995 a 2014. Os tipos de proposições legislativas analisadas são: Projeto de Lei (PL), Projeto de Emenda Constitucional (PEC), Projeto de Lei Complementar (PLP) e Medidas Provisórias (MP). No total são 42.441 proposições, incluindo proposições aprovadas, rejeitadas ou ainda em consideração.

Dessas 42.441 proposições, uma amostra aleatória de 15% foi selecionada para ser o training-set e foi codificada manualmente por um pesquisador treinados. A amostra aleatória simples se mostrou a mais apropriada de acordo com testes de Collingwood & Wilkerson (2012). O tamanho da amostra foi definido com base nos cálculos de custo-benefício de Collingwood & Wilkerson (2012). Isso porque é o menor tamanho da amostra para a maior precisão na classificação. O restante das proposições foi definido como o test-set, ou o grupo de proposições a ser classificado pelo algoritmo de aprendizado supervisionado. O algoritmo utilizado é o "support vector machine" (SVM). A escolha se deu por esse ser um algoritmo que exige menos memória e também porque gerou os melhores resultados de acordo com estimativa de Collingwood & Wilkerson (2012). A implementação da análise se deu no R com o uso do pacote RTextTools (Jurka *et al.*, 2013). A próxima seção apresenta os resultados preliminares.

4 Resultados Preliminares

Os resultados aqui apresentados são extremamente preliminares, uma vez que a aplicação da técnica ainda está sendo aprimorada para poder apresentar resultados finais. O primeiro passo para testar o algoritmo de classificação é dividir a amostra de proposições codificadas manualmente e usar uma metade para "ensinar" a máquina a classificar a outra metade. Esse procedimento é importante uma vez que para essa amostra nós sabemos a classificação "correta". Por isso se recomenda que esse importante passo seja feito, uma vez que para a

classificação de textos ainda sem classificação (limpos) não temos como saber se o algoritmo está acertando ou não. Realizamos a análise aqui apresentada da seguinte forma: das 5.700 proposições classificadas manualmente, dividimos o arquivo em 2.850 que foram usadas como o training-set e as outras 2850 como o test-set. Essa divisão não precisa ser exatamente a metade. Optamos por essa distribuição porque o algoritmo funciona melhor com maiores quantidades, independente da proporção.

Os resultados para esse teste estão na tabela 2. As estatísticas de interesse são Precision, Recall e F Score. Precision se refere à medida em que o algoritmo prevê corretamente a categoria da proposição ou "given that the machine guesses category k , what is the probability that the machine made the right guess" (Grimmer & Stewart, 2013, 14). Essa medida é calculada como o número de proposições corretamente classificados como categoria k , dividido pelo número total de proposições que a máquina classifica como categoria k . Note que essa análise só faz sentido como um teste em amostras já classificadas, já que em aplicações reais não saberemos quão preciso o algoritmo de fato é. Recall diz respeito à proporção de proposições numa categoria que o algoritmo prevê corretamente ou "given that a human coder labels a document as belonging to category k , what is the chance the machine identifies the document" (Grimmer & Stewart, 2013, 14). Essa estatística é calculada como o número de documentos corretamente classificados na categoria k , dividido pelo número de proposições classificadas manualmente como categoria k . O F Score é a média ponderada das duas estatísticas.

Tabela 2: Resultados - SVM

Precision	Recall	F Score
0.626	0.453	0.492

Os resultados alcançados com a amostra de proposições legislativas no Brasil em muito se aproxima à mesma realizada com dados do "Congressional Bills Project" (Collingwood & Wilkerson, 2012). A precisão é de 0.6, que é uma boa precisão, mas mostra que o algoritmo ainda erra muito na classificação. O Recall é de 0.45 e é mais baixo do que o valor encontrado para o mesmo teste com os dados americanos. Isso pode estar acontecendo por dois motivos: primeiro, a classificação manual pode estar com problemas e isso diminui o desempenho do algoritmo. Segundo, os textos das ementas devem ser melhor "limpados" para a análise, como a eliminação de palavras comuns, espaços, pontuação, dentre outros. Uma outra forma

de aumentar o desempenho da classificação automatizada é usar mais de um algoritmo e identificar os casos onde os algoritmos divergem. Esse raciocínio é o mesmo usado para a codificação manual, na qual o padrão é que dois codificadores classifiquem os mesmos dados e depois a concordância entre eles seja calculada. Como afirmado anteriormente, a análise aqui apresentada é ainda preliminar e esses serão os próximos passos da pesquisa.

Ainda que limitada, essa estratégia permite a classificação de grandes quantidades de texto. De forma preliminar, apresentamos alguns resultados utilizando essa metodologia. O gráfico 1 apresenta o percentual de proposições classificadas em cada uma das 20 categorias. Focamos a análise aqui nas categorias macro e não apresentamos resultados para as subcategorias, uma vez que não temos espaço para a discussão de resultados para mais de 200 subtópicos. A distribuição nas categorias dá uma noção geral de quais são os temas mais enfatizados na agenda legislativa brasileira.

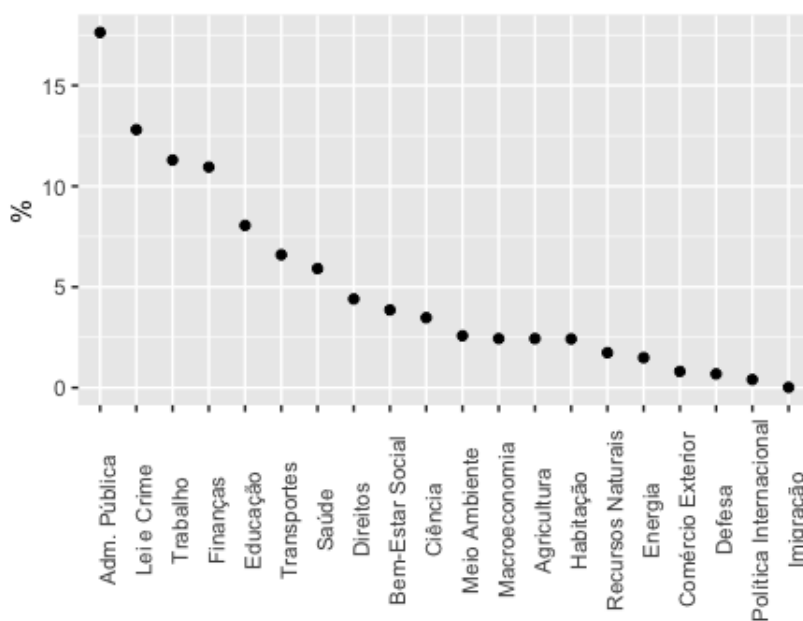


Figura 1: Saliência Temática Total (1995-2014)

A partir da figura 1 notamos que a maior ênfase na agenda legislativa no Brasil é em temas relacionados à Governo e Administração pública. Essa categoria engloba tópicos como a administração geral, que inclui muitos projetos de lei de solicitação de crédito adicional pelo Executivo, e também tópicos como funcionalismo público, nomeações e homenagens. O segundo tema mais enfatizado na agenda é a categoria Lei, Crime e Família. Essa categoria

inclui temas relacionados à justiça como instituições coercitivas, os códigos penal e civil e também questões relacionadas à definição e problemas de família e abuso de menores. Outros temas de grande ênfase são relacionados à relações trabalhistas, finanças e educação. De forma inesperada o tema de transportes aparece como um dos mais enfatizados pelos atores políticos brasileiros. Entre os temas menos enfatizados temos imigração, política internacional e defesa. Essa ênfase reduzida seria esperada para o caso brasileiro uma vez que o país não se insere em grandes conflitos e questões internacionais. Importante notar também a relevância de se construir medidas próprias para o contexto local, uma vez que há um certo consenso em torno da alta saliência desses temas em países europeus.

Como essas ênfases variam no tempo? A figura 2 apresenta a saliência temática por ano no período aqui analisado. Nota-se que há variação anual para alguns temas, mas não para outros. Na categoria trabalho há uma certa alternância entre anos nos quais a ênfase aumenta e anos em que a ênfase diminui. Já tópicos como imigração, comércio exterior, política internacional e recursos naturais a saliência dos temas se mantém estável nos 20 anos analisados.



Figura 2: Saliência Temática por Ano

As figuras 3 e 4 dividem as proposições legislativas de acordo com a autoria. A figura 3 foca nas proposições de origem do Executivo e a figura 4 apresenta as proposições de origem do Legislativo. A figura 3 apresenta a comparação na saliência dos diferentes temas por governo. A partir dela podemos notar as prioridades do Executivo e também como essas ênfases variam

de acordo com o governo em questão. Primeiramente, nota-se que as ênfases do Executivo se concentram em temas relacionados à administração pública, principalmente proposições sobre abertura de crédito, questões relacionados ao trabalho e também à justiça (lei, crime e família). Esse é o núcleo central de proposição do Executivo. Dos temas menos enfatizados, para além dos temas relacionados à política internacional e imigração, podemos destacar a relativa pequena ênfase do Executivo em temas classificados como bem-estar social, direitos e meio ambiente.

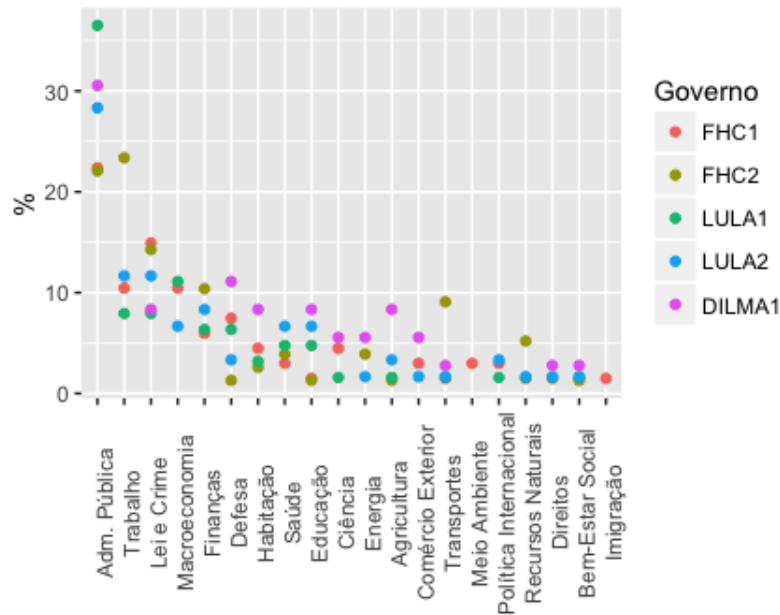


Figura 3: Saliência Temática - Executivo

Com relação à comparação entre diferentes governos, os cinco governos aqui analisados são os dois mandatos de Fernando Henrique Cardoso (FHC), os dois mandatos de Luiz Inácio Lula da Silva (LULA) e o primeiro mandato de Dilma Rousseff (DILMA). A distribuição dos percentuais mostra que em alguns temas a saliência se mantém estável durante os 5 governos. Esses são os temas de menor saliência durante todo o período como Imigração, política internacional e meio ambiente. Em outros temas há variação expressiva, como a forte ênfase em administração pública no primeiro mandato de Lula, a alta ênfase em trabalho e emprego no segundo governo FHC, a alta ênfase em defesa no primeiro mandato de Dilma, a maior ênfase em agricultura no primeiro mandato de Dilma e a alta ênfase em transportes no segundo mandato de FHC.

Notar que essas variações entre governo podem existir também como uma dependência

temporal. Isso porque quando há uma "saturação" de iniciativas numa legislatura o tema pode ser deixado de lado na próxima. Notar também que a lógica de proposição legislativa do Executivo pode ser diferente da lógica dos parlamentares. Isso porque os poderes de agenda dos parlamentares são limitados e a proposição legislativa serve mais como uma medida de position-taking, evidenciando mais claramente as ênfases temáticas ou a saliência como o buscado nesse artigo. No caso da proposição pelo Executivo a lógica pode ser muita mais de efetividade do que de tomada de posição. Isso porque a não votação ou não aprovação de medidas de iniciativa do Executivo podem ser interpretadas como fracasso da atuação do presidente. Então, quando o Executivo faz uma proposição é para essa ser aprovada e não para sinalizar algo ou para tomar posição. A figura 4 apresenta a distribuição de ênfases com foco nas proposições dos parlamentares.

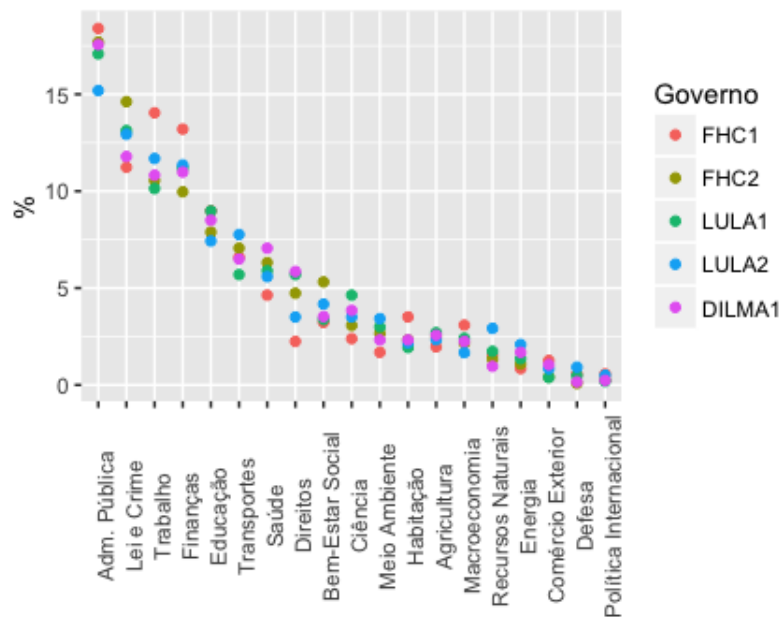


Figura 4: Saliência Temática - Legislativo

Nas ênfases temáticas dos parlamentares podemos notar menor concentração do que nas ênfases do Executivo. A ordem de prioridades também muda. Para os parlamentares, os temas mais salientes são administração pública, lei e crime, trabalho, finanças e educação. Os menos salientes são, novamente, política internacional, defesa, comércio exterior, energia e recursos naturais. Algo que pode ser observado também a partir da figura é que há maior homogeneidade na comparação entre os governos. Isso porque nas proposições do Executivo

a mudança na presidência afeta diretamente a agenda do Executivo. No caso do Legislativo, por ser uma agenda descentralizada, as mudanças de composição são menos extremas quando comparado à mudança de um presidente.

A figura 5 apresenta a atuação dos partidos políticos em termos de suas ênfases temáticas. Cada gráfico mostra o percentual de proposições em cada área temática analisada. O objetivo é identificar se há ênfases específicas dos diferentes partidos.

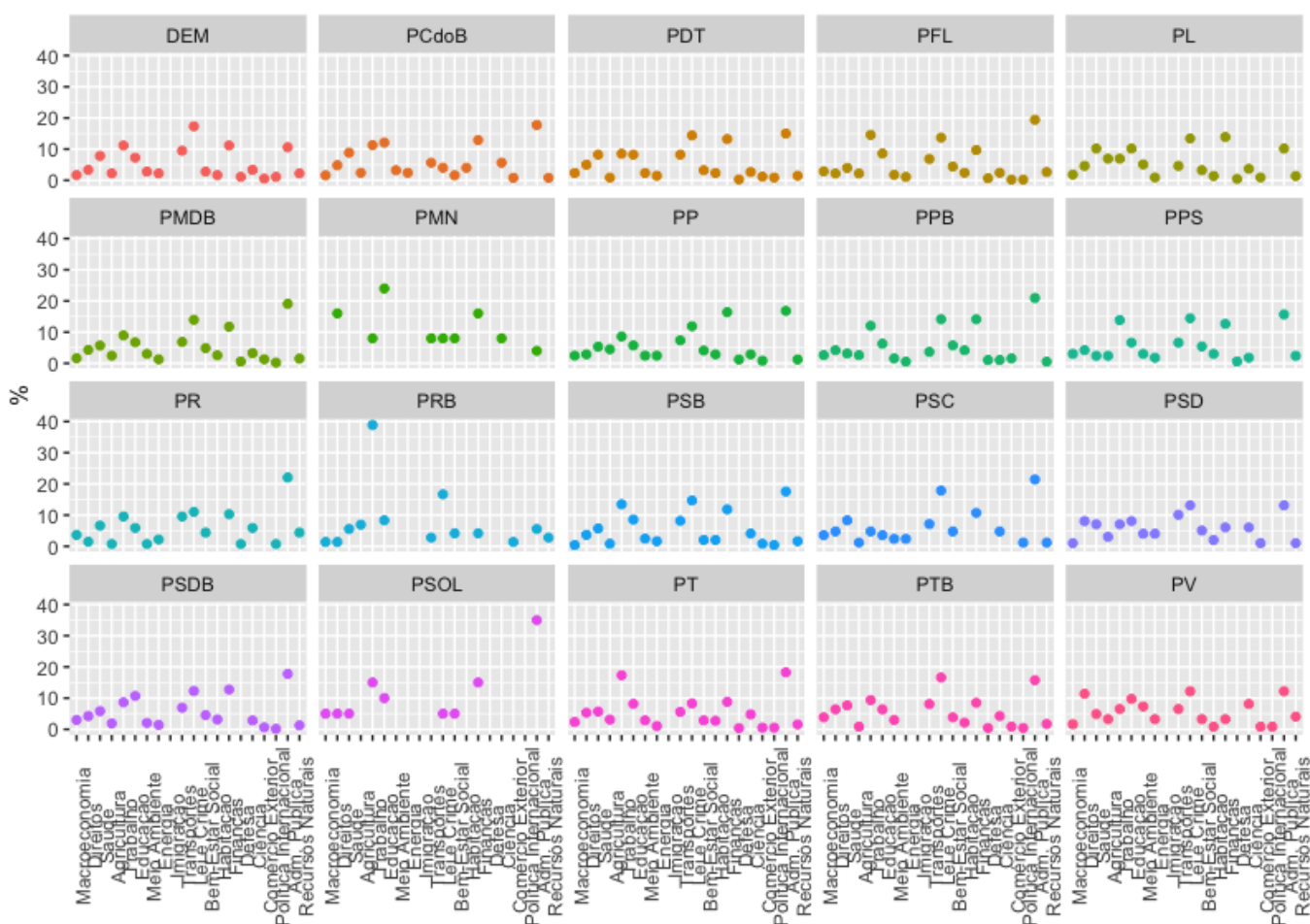


Figura 5: Saliência Temática - Partidos

A partir da figura 5, em geral podemos notar uma certa homogeneidade na distribuição das ênfases temáticas. Isso ocorre sobretudo devido à heterogeneidade na composição das bancadas partidárias, que nem sempre tem seus membros se preocupando ou defendendo os mesmos temas. Contudo, alguns pontos se destacam em alguns partidos. Alguns desses exemplos são a ênfase do PRB em tópicos relacionados ao trabalho e emprego, a ênfase do

PSOL em temas de administração pública, a ênfase do PSC em temas relacionados a lei, crime e família e a ênfase do PT em temas relacionados ao trabalho.

Por último, apresentamos a distribuição das ênfase de acordo com os partidos políticos. Essa figura se relaciona diretamente à teoria da saliência que prevê a existência de "monopólios" temáticos, ou seja, o domínio de um determinado tema por um partido específico. Isso ocorreria como forma de se diferenciar e adquirir vantagem na competição política. A existência de monopólios temáticos seria observada através da alta concentração da proposição em determinado tema por partidos específicos. Um exemplo de monopólio que esperaríamos encontrar seria a concentração da ênfase temática de um partido verde na área de meio ambiente. A figura 6 apresenta essa distribuição para o caso brasileiro.

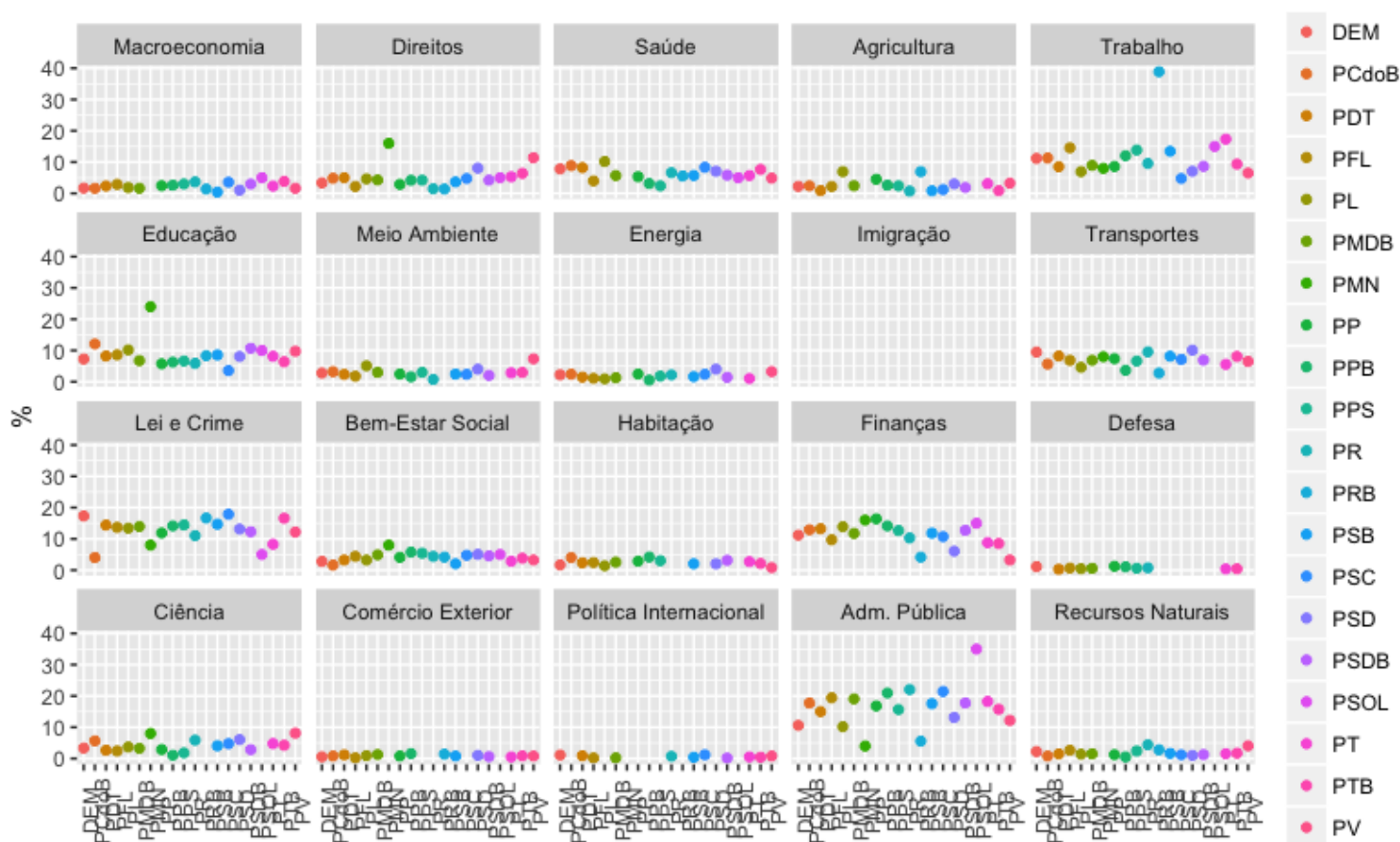


Figura 6: Saliência Temática - Monopólios

A partir da figura 6 identificamos que não há monopólios claros na atuação dos parlamentares brasileiros. Algumas pequenas exceções seriam a ênfase do PSOL em temas de administração pública, a ênfase do PMN em temas de educação e direitos e a ênfase do PPB

em temas de trabalho. Notar que essas exceções em termos de concentração da atenção num determinado tema se dá especificamente em partidos pequenos, onde a homogeneidade seria maior devido a um menor número de parlamentares. Por essa razão, acreditamos que os resultados aqui apresentados não oferecem evidência de existência de monopólios temáticos e também mostra uma heterogeneidade partidária quanto à saliência temática, sendo algo muito mais relacionado ao parlamentar individual.

Não buscamos aqui explicar essas variações de saliência ou porque alguns temas são mais enfatizados por um determinado partido ou num determinado período de tempo. Temos por objetivo apenas que os dados aqui apresentado possam ser refinados e sirvam de base para o teste de tantos argumentos ainda em aberto na literatura no Brasil.

5 Conclusão

Esse trabalho teve como objetivo implementar uma metodologia para classificação das proposições legislativas no Brasil e assim construir uma medida de saliência temática que possa ser aplicada para dados num longo período de tempo e para diversos atores políticos. Apresentamos o esquema de codificação do "Policy Agendas Project" e utilizamos a abordagem de aprendizado supervisionado para utilizar uma amostra de proposições codificadas para ensinar um algoritmo para classificar o total de proposições legislativas no Brasil no período de 1995 a 2014.

Os resultados aqui apresentados são ainda preliminares e a metodologia, assim como os dados, precisam ser refinados. Contudo, para além de oferecer respostas, buscamos aqui simplesmente abrir o debate para a possibilidade de utilização dessa metodologia para construção de indicadores de ênfases temáticas que possam ser utilizados para responder questões relativas a temas de alta importância, desde comportamento partidário, passando por relações Executivo-Legislativo, até a atuação de grupos de interesse no Congresso Nacional.

Referências

- Baumgartner, Frank R, & Jones, Bryan D. 2010. *Agendas and instability in American politics*. University of Chicago Press.
- Baumgartner, Frank R, Jones, Bryan D, & Wilkerson, John. 2011. Comparative studies of policy dynamics. *Comparative Political Studies*, 00.
- Budge, Ian. 2001a. *Mapping policy preferences: estimates for parties, electors, and governments, 1945-1998*. Vol. 1. Oxford University Press on Demand.
- Budge, Ian. 2001b. Validating the manifesto research group approach. *In: Laver, Michael (ed), Estimating the policy position of political actors*. Oxford: Psychology Press.
- Collingwood, Loren, & Wilkerson, John. 2012. Tradeoffs in accuracy and efficiency in supervised learning methods. *Journal of Information Technology & Politics*, **9**(3), 298–318.
- Figueiredo, Argelina, & Limongi, Fernando. 2001. *Executivo e Legislativo na Nova Ordem Constitucional*. Rio de Janeiro: FGV.
- Green-Pedersen, Christoffer, & Walgrave, Stefaan. 2014. *Agenda setting, policies, and political systems: a comparative approach*. University of Chicago Press.
- Grimmer, Justin. 2013. *Representational Style in Congress: What Legislators Say and Why It Matters*. Cambridge University Press.
- Grimmer, Justin, & Stewart, Brandon M. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, mps028.
- Hillard, Dustin, Purpura, Stephen, & Wilkerson, John. 2008. Computer-assisted topic classification for mixed-methods social science research. *Journal of Information Technology & Politics*, **4**(4), 31–46.
- Humphreys, Macartan, & Garry, John. 2000. Thinking about salience. *Harvard University. Unpublished manuscript*.
- Jurka, Timothy P, Collingwood, Loren, Boydston, Amber E, Grossman, Emiliano, & van Atteveldt, Wouter. 2013. RTextTools: A supervised learning package for text classification. *The R Journal*, **5**(1), 6–12.

- Kingdon, John W. 2003. *Agendas, alternatives, and public policies*. Longman Pub Group.
- Krippendorff, Klaus. 2004. *Content analysis: An introduction to its methodology*. Sage.
- Lasswell, Harold Dwight. 1950. *Politics: Who gets what, when, how*. P. Smith New York.
- Laver, Michael. 2001. Position and salience in the policies of political actors. *In*: Laver, Michael (ed), *Estimating the policy position of political actors*. Oxford: Psychology Press.
- Laver, Michael, & Hunt, Ben. 1992. *Policy and Party Competition*. New York: Routledge.
- Laver, Michael, & Shepsle, Kenneth. 1994. *Cabinet Ministers and Parliamentary Government*. Cambridge: Cambridge University Press.
- Laver, Michael, *et al.* . 2003a. *Estimating the policy position of political actors*. Routledge.
- Laver, Michael, Benoit, Kenneth, & Garry, John. 2003b. Extracting policy positions from political texts using words as data. *American Political Science Review*, **97**(02), 311–331.
- Moreira, Davi. 2016. *Com a palavra os nobres deputados frequência e ênfase temática dos discursos dos parlamentares brasileiros*.
- Neuendorf, Kimberly A. 2002. *The content analysis guidebook*. Sage.
- Purpura, Stephen, & Hillard, Dustin. 2006. Automated classification of congressional legislation. *Pages 219–225 of: Proceedings of the 2006 international conference on Digital government research*. Digital Government Society of North America.
- Quinn, Kevin M, Monroe, Burt L, Colaresi, Michael, Crespin, Michael H, & Radev, Dragomir R. 2010. How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, **54**(1), 209–228.
- Robertson, David Bruce. 1976. *A theory of party competition*. John Wiley & Sons.
- Tarouco, Gabriela. 2011. Brazilian parties according to their manifestos: political identity and programmatic emphases.
- Tarouco, Gabriela, & Madeira, Rafael Machado. 2013. Esquerda e direita no sistema partidário brasileiro: análise de conteúdo de documentos programáticos. *Revista Debates*, **7**(2), 93–114.

- Volkens, Andrea. 2001. Manifesto research since 1979: From reliability to validity. *In*: Laver, Michael (ed), *Estimating the policy position of political actors*. Oxford: Psychology Press.
- Warwick, Paul, & Druckman, James. 2001. Portfolio Saliency and Proportionality of Payoffs in Coalition Governments. *British Journal of Political Science*, **31**1.
- Warwick, Paul V, & Druckman, James N. 2006. The portfolio allocation paradox: An investigation into the nature of a very strong but puzzling relationship. *European Journal of Political Research*, **45**(4), 635–665.
- Wlezien, Christopher. 2005. On the saliency of political issues: The problem with ‘most important problem’. *Electoral Studies*, **24**(4), 555–579.