

Análise contrastiva das expressões binominais em *corpora* anotados portugueses e franceses

Contrastive analysis of compound expressions in Portuguese and French annotated *corpora*

FERNANDO MARTINHO*

PALAVRAS-CHAVE: Sintaxe, Léxico, Modificadores, Adjetivos, Compostos, Semântica intensional.

KEYWORDS: Syntax, Lexicon, Adjectival modifiers, Attributive adjectives, Compounds, Intensional semantics.

Introdução

O estudo do léxico revela que existe por vezes conflito entre a natureza e a função de alguns itens lexicais. O presente trabalho contrastivo, baseado em dados da linguística de corpus, propõe-se refletir sobre um conjunto de termos das línguas portuguesa e francesa, incluídos por defeito na classe dos nomes, mas interpretados, em contextos específicos, como adjetivos. No caso da língua portuguesa, retomamos aqui parte dos dados extraídos do corpus informatizado *CETEM Público*¹. No caso da língua francesa, propomos tratar dados extraídos do corpus *frWaC*² e cruzá-los com os dados do português.

Na sua qualidade de modificadores, os adjetivos atributivos são classicamente vistos como unidades lexicais superficialmente adjacentes ao núcleo lexical

* Universidade de Aveiro. Centro de Línguas, Literaturas e Culturas (CLLC).

¹ <http://www.linguateca.pt/cetempublico/>. O *CETEM Público* (Corpus de Extratos de Textos Eletrónicos MCT/Público) é um corpus de aproximadamente 180 milhões de palavras em português europeu, codificado com IMS Open CWB e criado pelo projeto Processamento computacional do português (projeto que deu origem à Linguateca) após a assinatura de um protocolo entre o Ministério da Ciência e da Tecnologia (MCT) português e o jornal público em abril de 2000.

² <https://bellatrix.sslmit.unibo.it/noske/public/#dashboard?corpname=frwac1>. *FrWaC* é um corpus de 1,613 mil milhões de *token* (1,278 mil milhões de palavras) construído a partir da Web, limitando o rastejar ao domínio .fr e usando palavras de média frequência do *Le Monde Diplomatique* e listas básicas de vocabulário francês como entradas. O corpus foi marcado por POS e lematizado com *Tree Tagger*.

N, partilhando com o mesmo os seus traços flexionais. No entanto, nas línguas referidas, o estudo dos *corpora* citados permite identificar casos de modificação nominal, que implicam, da mesma forma, adjacência entre dois N, não estando, contudo, o primeiro (N1) e o segundo (N2) necessariamente ligados por complementação sintática, relações argumentais ou aposição. Nesta configuração particular, veremos que N2 pode ser interpretado como modificador de N1. O presente estudo refere-se, em concreto, a um subconjunto de sequências N1N2 à margem dos processos clássicos de composição (que, por sua vez, têm uma frequência de corpus muito elevada), o qual corresponde, como será defendido, a uma estrutura modificada em que N2 é um adjetivo. Concretamente, nestas expressões binomiais modificadas, N2 corresponde a um adjetivo intensional com função relacional ou superlativa.

1. Binomiais e compostos

Apresentamos, para começar, os *corpora* referidos e algumas concordâncias extraídas, e retomamos depois algumas análises relevantes sobre expressões binomiais.

1.1. *Corpus* de trabalho

Os exemplos em (1.) correspondem a expressões binomiais N1N2 resultando de concordâncias extraídas aleatoriamente da base de dados CETEM³:

(1.) Expressões binomiais N1N2 em Português (*ortografia original*)
projecto-piloto, contrato-promessa, contrato-programa, projecto-lei, medida-padrão, clube-satélite, estado-membro, episódio-chave, palavra-chave, ministro-presidente, posto-chave, versão-base, emissor pirata, equipa sensação, estado-fantoches, trabalhador-estudante,

³ Dados extraídos com base na sintaxe de pesquisa [pos="N"][pos="N], em que “pos” é a instrução de “posição” e “N” a categoria “nome”. Note-se que o corpus CETEM contém um número considerável de N1N2 hifenizados. Acesso: <https://www.linguateca.pt/CETEMPublico/>

O corpus em (2.) ilustra, em francês, o mesmo tipo de sequências N1N2, aleatoriamente selecionadas. Cada uma corresponde a uma expressão binominal e resulta de concordâncias extraídas via CQL⁴ do corpus FrWac⁵ uma vez depuradas de falsos positivos⁶:

(2.) Expressões binominais N1N2 em Francês⁷:

image système, performance record, responsable communication, pièce moto, service achat, technologie objet, coloration métier, prix monstre, gris acier, effet lumeau, mots clés, assurance moto, édition papier, association relais,

1.2. Modelos teóricos

No campo da morfossintaxe e do estudo lexical das composições, em particular nas línguas românicas, vários autores abordaram as estruturas adjacentes do tipo N1N2 e procuraram opô-las a construções envolvendo outras categorias lexicais contíguas (como NA ou VN). Assim, é geralmente assumido que as sequências N1N2 se distinguem dos casos de aposição, não estando N1 e N2 separados por vírgula ou pausa, dos casos de predicação, pois não existe entre N1 e N2 nenhuma cópula, e também dos casos de complementação, pois nenhuma preposição ou conjunção é inserida entre N1 e N2. Em contrapartida, a adjacência entre os termos aponta para mecanismos de composição ou de

⁴ O *Corpus Query Language* (CQL) é um código especial de consulta usado no *Sketch Engine* (<http://www.sketchengine.eu/>) para procurar padrões gramaticais ou lexicais complexos ou para usar critérios de pesquisa que não podem ser definidos usando a interface padrão do utilizador. A expressão regular principal do CQL, utilizada para esta consulta de corpus, foi: [tag="NOM"][tag="NOM"]

⁵ O corpus estava aqui restrito ao subconjunto *FrWac Small* (uma variante de mil milhões de palavras). O acesso à página principal é: <https://bellatrix.sslmit.unibo.it/noske/public/#dashboard?corpname=frwac1>

⁶ A consulta inicial do CQL devolveu cerca de 459 000 concordâncias em 22 000 páginas! Embora a opção *Sentence View* tenha ajudado pela primeira vez na filtragem de operações, decidimos extrair dados de apenas uma em cada dez páginas (implicando cerca de 2 000 páginas), o que resultou em 40 000 concordâncias. A filtragem referida foi aplicada a estas concordâncias posteriores.

⁷ Em francês, um número muito reduzido de binominais usa hífen entre N1 e N2, casos de: “*ministre-candidat, achat-vente*”, etc. Note-se, ainda, que estas sequências podem, para além do hífen, ser também transcritas com aspas ou maiúscula (no caso de nomes próprios).

modificação sintática ou morfológica. Neste último caso, têm sido propostas designações como ‘substantivo adjetivado’ (Béchade, 1986 apud Noailly, 1990) ou ‘nome adjetivo’ (Martinho, 2007), para denotar estas estruturas indecisas que, na verdade, carecem de classificação e de apelação rigorosas⁸.

1.2.1. Composição e modificação

A definição dos compostos e a distinção entre composição e modificação é uma questão recorrente, abordada tanto no âmbito da gramática normativa como dentro de diversos quadros teóricos (Bartning 1976, Warren 1984, Vogel e Comrie 2000, Olsen 2001, Booij 2007).

Em termos de norma gramatical, o composto N1N2 tem sido considerado uma construção depreciativa, por referência à classificação léxico-semântica herdada da gramática tradicional, cuja distinção aristotélica entre o “nome substantivo”, que denota as substâncias e “nome adjetivo”, que denota as qualidades, continua historicamente vincada⁹. Logo, a tradição gramatical francesa, por exemplo, Grevisse e Goosse (1961), defende que a função de modificador atributivo – ou “*epithète*”¹⁰ – pertence em exclusivo ao adjetivo, pelo que sequências como “*roman fleuve*” são compostas de nomes substantivos. Barbaud (1971, p. 102) argumenta que um “*composé binominal*” associando dois nomes justapostos, como “*centre ville*”, deve na verdade ser pensado como uma expressão N1(*de*)N2 em que o segundo pode ser considerado um complemento dativo ou genitivo, depois de elipse da preposição intermédia “*de*”. Considerando “marginais” estas composições, Barbaud reconhece, contudo, que, nas construções N1N2, ditas de “*complementação direta*”, a supressão de “*de*” é cada vez mais usada na língua moderna¹¹. De igual modo, relativamente à norma gramatical do português, Cunha e Cintra (1984) encaram também como casos de “composição por justaposição” sequências de dois substantivos como “*escola-modelo*” ou “*porco-espinho*”, sendo N1 “*o elemento determinado, que contém a*

⁸ Sobre uma proposta de distinção entre “nomes adjetivos” e “adjetivos nomes” (Martinho, 2007).

⁹ Cunha e Cintra (1984, p. 203) referem a “duvidosa vernaculidade” dos substantivos usados como adjetivos de cor (“*verde-garrafa*”)

¹⁰ Da mesma forma, a função predicativa do adjetivo é qualificada de “*attribut*” (atributo).

¹¹ Barbaud (1971, p. 76). Na mesma perspetiva, Rohrer (1967 apud Noailly, 1990) considera agramaticais as construções N1N2.

ideia geral” e N2 “*o determinante, que encerra a noção particular*”¹² (1984, p. 107).

No âmbito da análise linguística do português, as sequências N1N2 têm sido pensadas em termos de composição morfológica ou sintática (ou ambas)¹³. Assim, Villalva (2003, p. 980) refere globalmente os “compostos morfossintáticos”, os quais incluem em particular a estrutura N1N2 e remetem para duas grandes categorias: (i) as *estruturas de adjunção*, como “bomba-relógio”, que correspondem a uma expressão nominal “de núcleo inicial” constituída por dois N, sendo N2 um “modificador nominal” (habitualmente) não flexionado; e (ii) as estruturas de *conjunção*, como “surdo-mudo”, cujo núcleo não pode ser isolado e N2 pode flexionar. Rio-Torto (2013, p. 21) distingue (i) os compostos por coordenação, como “porco-espinho”, que permitem uma denotação por intersecção entre os dois N; e (ii) os compostos por subordinação, como “turma-piloto”, que não implicam uma denotação interseccional e se baseiam numa relação “Determinado-Determinante”, na qual o elemento determinante (N2) é um substantivo¹⁴. Para esta autora, os “nomes em aposição” N1N2 correspondem a um tipo específico de composição em que se destacam duas constantes. Por um lado, a construção N1N2 é um composto de natureza morfológica ou morfossintática e pode ser declinada, em função dos casos, em coordenativa (“trabalhador-estudante”), subordinativa (“salva-vidas”) ou atributiva (“palavra-chave”). Por outro lado, em compostos atributivos, N2 funciona como um modificador de N1, embora a sua flexão seja inconstante.

Relativamente ao francês, Noailly (1990) defende a proposta segundo a qual N2 seria um adjetivo. Longe de ser marginal, este uso inesperado do nome é amplamente registado no francês contemporâneo, onde compete – com sucesso crescente – com as estruturas de modificação adjetival clássicas. Noailly distingue nos “*substantifs épithètes*” dois tipos de interpretações. Destaca, em primeiro lugar, a “qualificação”, função habitualmente desempenhada por um adjetivo qualificativo e eventualmente por um nome comum. Por seu lado, a função de “*identificação*” está associada a um nome (comum ou próprio) ou

¹² Os autores observam, acerca do exemplo “*mãe-pátria*”, que a ordem *Determinante-Determinado* (N2N1) também pode existir em português.

¹³ Nessa perspetiva, as eventuais relações gramaticais entre N1 e N2 não seriam exclusivas do composto, mas presumivelmente comuns às construções sintáticas em geral.

¹⁴ (Rio-Torto, 2013, pp. 21-23)

a um adjetivo relacional¹⁵. Vejamos exemplos de *qualificação* e *identificação* em (3.) e (4.):

- (3.) a. Un livre événement
b. Une émission phare
- (4.) a. Un plateau repas
b. Un chèque vacances

Para Noailly (1990, p. 38), N2 é “qualificativo” a partir do momento em que N1N2 pode ser parafraseado por “um N1 que é um N2», isto é, em que N2 equivale a uma relativa restritiva. Assim, um “*livre-événement*” é “um livro (N1) que é um acontecimento (N2)”¹⁶ (Noailly, 1990, p. 36). A autora postula também uma equivalência relativa entre sequências N1N2 modificadas, como “*film-scandale*” e sequências NA clássicas, como “*film scandaleux*”: em ambos os casos, o referente (N1) é modificado por meio de um “*qualificativo*”; a paráfrase proposta permite alternar entre um nome, que aceita então um artigo (o que comprova a sua natureza nominal – “um filme que é *um* escândalo”), e um adjetivo derivado via sufixação (o que comprova a sua natureza adjetival – “um filme que é *escandaloso*”¹⁷). Quanto a N2 de “identificação”, a autora focaliza-se em casos como “*un problème cheveux*”, “*les sœurs courage*”, para invocar constantes subjacentes, entre as quais a ausência de concordância entre N1 e N2, que afasta N2 dos adjetivos, a ausência sistemática de hifenização (ao contrário dos N1N2 de “qualificação”), a ausência de grau ou sufixo em N2 e, por último, a incompatibilidade de N2 com uma eventual posição predicativa¹⁸.

A análise linguística das estruturas N1N2 pode ser alargada a outras línguas. Numa perspetiva próxima, Montermini (2006) propõe, em italiano, que os “compostos” N1N2 resultam de uma fusão entre dois nomes numa relação de *modificação*. O autor sugere que N1N2 deve ser interpretado como *hipónimo*

¹⁵ Havendo complementação, estas sequências podem também ser parafraseadas por meio de uma construção preposicional ou relativa, como por exemplo: “*le facteur risque*” / “*le facteur qui est un risque*” (Noailly, p. 146).

¹⁶ A autora critica Grevisse e Goosse (1961), que confundem, segundo ela, “*qualificação*” e “*coordenação*”.

¹⁷ Este critério é claramente fragilizado por casos como o N1N2 “*prix monstres*”, que apenas pode ser parafraseado em “*des prix qui sont monstres*” (*Idem*, p. 39).

¹⁸ Exemplo de N1N2 agramatical: “*un plateau très repas*”. Estas restrições levam a autora, crucialmente, a afastar N2 dos adjetivos qualificativos e a aproximá-lo dos adjetivos relacionais. Para uma proposta de análise de N2 como adjetivo relacional, cf. Martinho (2013).

de N1, pelo que adianta, à semelhança de Noailly (1990), uma paráfrase restritiva de N1N2: “um N1 que é ao mesmo tempo um N2”, apresentando a favor dessa análise exemplos como “*viaggio premio, pausa pranzo, angolo cottura*”¹⁹. Dressler (2006), numa tentativa de apurar a definição dos “*compounds*” em inglês, sugere que, se é difícil atribuir-lhes um formato consensual, em contrapartida a sua semântica parece ser facilmente mapeada. O autor focaliza-se na análise extensional dos compostos, mais precisamente nas suas propriedades prototípicas, e no facto de os mesmos serem constituídos de palavras básicas não derivadas “*free forms*”, excluindo afixos, clíticos, preposições²⁰. Analisando a produtividade morfológica dos “*compound nouns*”, o autor nota a sua tendência em se comportarem como estruturas endocêntricas de núcleo final, ou seja, do tipo *Determinante-Determinado*²¹, como “*miracle cure*” ou “*high-school*”. Por último, o autor refere que os compostos são ilhas anafóricas (“*anaphoric islands*”), ou seja, a sua estrutura interna é opaca a operações sintáticas: assim, em “*truck driver*”, o N2 “*truck*” não é acessível a leituras referenciais, como o atesta a agramaticalidade da frase seguinte (em que o referente de “*them*” seria “*truck*”): * “*Truck drivers fill them up*”.

Em síntese, com base nos modelos anteriores, podemos associar a natureza e a estrutura dos compostos a algumas constantes: (i) os mesmos integram sistematicamente duas ocorrências da categoria lexical N; (ii) N1 e N2 verificam entre si uma relação linguística implícita (isto é, não formal) em que N2 determina e N1 é determinado; (iii) a estrutura resultante (nas línguas Românicas) é um hipónimo de N1 e pertence à categoria lexical de N1. Uma das questões em aberto consiste em determinar a natureza da relação linguística verificada entre N1 e N2, a qual pode ser de tipo sintático, morfológico e/ou semântico.

1.2.2. Composicionalidade

Um princípio relevante que caracteriza o conhecimento linguístico implícito de um falante (em particular o seu “léxico mental”²²), e que é solicitado em particular no processamento destas sequências lexicais ambíguas, é o princípio

¹⁹ Lit.: “viagem prenda, pausa almoço, ângulo cozedura” Montermini (2006, p. 2).

²⁰ Como seria o caso, por exemplo, de “*biblio-therapy*” (Dressler, 2006, p. 25).

²¹ Neste caso, em inglês.

²² Libben e Jarema (2006).

de composicionalidade²³. Em estruturas lexicais complexas, Veloso e Martins (2011) apontam para que “a computação do significado” se baseia nesse “princípio, que permite calcular (e torna predizível) o significado partilhado por toda a palavra através da identificação, em separado, do significado de cada *um* dos seus constituintes básicos” (Veloso e Martins, 2011, p. 50). Os autores retomam a definição de Villalva (2008), formulada nos seguintes termos²⁴: “Composicionalidade é um princípio estabelecido para a semântica, segundo o qual o significado de um todo é uma função do significado das partes” (Villalva, 2008, p. 22).

Este princípio aplica-se precisamente no caso das sequências N1N2, na medida em que a interpretação de expressões como “equipa sensação” ou “*bilan langage*” procede do significado de N1, mas também da relação de modificação estabelecida entre N1 e N2²⁵. O critério de composicionalidade permite, pois, estabelecer uma distinção pertinente e objetiva entre os binominais modificados N1N2, de baixo valor composicional, e os casos de composição (não modificada), de interpretação marcadamente composicional²⁶. Assim, segundo Veloso e Martins²⁷ o Princípio de Composicionalidade contribui, pelo menos em parte, para a explicação de numerosos casos de inovação lexical na língua, facilitando a entrada de palavras como [as EBM] no léxico português” (2011, p. 563).

Embora estes autores classifiquem globalmente os N1N2 como compostos, hesitam sobre os critérios de aplicação dos mecanismos de composição e o grau de composicionalidade. Um dos testes aplicáveis aponta para a morfologia interna de N1N2, visto que a flexão de N2, nestas estruturas, parece ser bastante

²³ O conceito remonta, segundo Veloso e Martins (2011), a Frege (1892).

²⁴ A autora alarga o conceito de composicionalidade também à morfologia.

²⁵ Sequências como “equipa sensação” ou “*bilan langage*” destacam-se também pelo facto de constituírem exemplos de sobregeração, ou seja, não estão ainda dicionarizadas. Como tal, a elevada produtividade destas expressões seria explicada à luz do princípio de composicionalidade (Veloso e Martins, 2011, p. 562).

²⁶ Observe-se, contudo, que parte dos compostos (por exemplo predicções do tipo “tira-nódoas”, etc.) equivalem a expressões não-composicionais, próximas de estruturas lexicalizadas.

²⁷ Os autores dão o exemplo de “governo sombra” e referem que se trata de “neologismos relativamente recentes que combinam palavras ou unidades morfológicas pré-existentes e disponíveis no léxico partilhado por um número supostamente substancial de falantes” (Veloso e Martins, 2011, p. 563).

inconsistente²⁸. Assim, se N2 mantém sempre o seu *gênero* (independentemente de N1), como em “a turma piloto” ou “*la mère courage*”, N2 pode ter ou não o número de N1 (“as turmas piloto”, “*les responsables clés*”). A composicionalidade da interpretação é um critério a considerar na medida em que a denotação de um composto não composicional não é transparente, ou seja, a interpretação implica uma dissociação referencial entre N1 e N2 – uma “turma piloto” é uma “turma”, mas não é um “piloto” –, o que não é o caso dos compostos composicionais (um “trabalhador estudante” é ao mesmo tempo um “estudante” e um “trabalhador”). Devido à sua inclinação para a lexicalização, as expressões N1N2 modificadas não são, pois, transparentes em termos composicionais²⁹. Nessa mesma perspetiva, Dressler (2006) explica que o significado de “*high school*”, por exemplo, é obviamente composicional e, portanto, transparente, mas o cognato “*high-school*” é opaco (logo, não composicional). Ainda assim, nota o autor, “*high-school*” pode ser considerado razoavelmente transparente, porque, se o significado de “*school*” é de facto extensional, a motivação semântica de “*high*” também parece clara. Podemos, pois, definir o significado de N1N2 como sendo um subconjunto do conjunto de significados construídos com base na combinação dos significados de N1 e N2, estando, contudo, a natureza nominal ou adjetival de N2 na origem de subconjuntos distintos.

Os modelos aqui sumariamente apresentados, que tentam atribuir uma classificação consistente a itens lexicais inconstantes, levantam certamente diversas questões sobre a sua aplicabilidade, em especial pela sua hesitação entre remeter explicações para o léxico, a morfologia ou a sintaxe, e pelo número consequente de exceções às regras projetadas³⁰. Esta questão tipológica e terminológica pode ser no entanto, como vamos ver, reduzida à proposta seguinte: as sequências N1N2 aqui consideradas não são nem lexicais nem

²⁸ Existe algum consenso sobre o facto de, nos N1N2 modificados, N2 ser flexionado. Contudo, como mostram os exemplos extraídos dos corpora, em muitos casos apenas N1 flexiona. Cf. Villoing (2012) ou Rio-Torto (2013).

²⁹ Acerca dos processos de lexicalização, Villalva (2008) adianta: “Quanto à lexicalização, trata-se de um processo de perda da composicionalidade, que actua de forma aleatória e imprevisível, sempre que pelo menos um dos constituintes morfológicos sofreu alterações semânticas ou formais ou é desconhecido para os falantes. A lexicalização pode, pois, afectar a interpretação da palavra, a sua forma, as suas propriedades gramaticais ou uma conjugação destes factores” (Villalva, 2008, p. 23).

³⁰ Outra questão é que estes estudos juntam dados e interpretações muito diferentes e, em última análise, dão pouquíssima atenção às sequências N1N2 propriamente ditas, privilegiando tipos de sequências mais frequentes (VN, AN...).

morfológicas, mas sintáticas. Nesta perspetiva, defendemos que um N2 inserido de forma não forçada (isto é, não lexicalizada) ao contacto de um N1 desprovido de função temática corresponde a um modificador.³¹ Propomos, assim, que uma expressão como “*visite éclair*” (visita relâmpago), ao contrário, por exemplo, de “*ministre-candidat*”, não é um composto, mas uma *expressão binominal modificada* (EBM), opaca e de natureza não-composicional, que, em termos sintáticos, corresponde a um SN de núcleo N1 e adjunto sintático N2; na estrutura interna deste SN, N2 equivale semântica e sintaticamente a um modificador adjetival.

2. Expressões binominais

As expressões N1N2 ilustram uma característica pertinente na descrição da norma gramatical e do conhecimento lexical – a da indefinição das fronteiras entre nomes e adjetivos –, bem como representam uma fonte óbvia de enriquecimento lexical e criação neológica, o que provavelmente explica que sejam hoje tão comuns, pois permitem uma extensão do léxico “de baixo custo”, isto é, com base em operações simples preexistentes (Libben, 2006). De forma geral, estas expressões, pelo facto de contornarem as estruturas de complementação ou predicação regulares, procuram desenvolver uma estratégia de comunicação distinta das construções sintáticas habituais, pelo que a sua existência poderá eventualmente ser motivada por imperativos específicos da comunicação humana³². Além disso, como observam Guevarra e Scalise (2009, p. 101), as expressões binominais não deixam de estar associadas, em termos tipológicos, a contextos não composicionais como as lexicalizações, colocações e idiomatismos, logo, estas estruturas destacam-se, da mesma forma, por comprimirem ou compactarem mais eficazmente o significante linguístico, reduzindo a uma simples estrutura N1N2 as informações a comunicar. Crucialmente, essa compressão dos dados é tornada possível pelo recurso à metaforização: processos

³¹ Na verdade, com base nos *corpora* referidos, veremos que as expressões N1N2 correspondem de forma maioritária a um conjunto de estruturas heterogéneas que não se relacionam com expressões modificadas. Nestes *corpora*, os binominais modificados representam uma pequena percentagem (15-20%) nas concordâncias de resposta às instruções [tag=“NOM”] [tag=“NOM”] ou [pos=“N”][pos=“N”].

³² Comrie (2003) e a relação que estabelece entre imperativos de comunicação e universais da linguagem.

de motivação metafórica permitem, nestas estruturas, a flexibilidade necessária para estabelecer relações de modificação entre N1 e N2 (Guevarra e Scalise, 2009, p. 104). Casos como “código fonte” ou “*prix monstre*” são expressões representativas dessa capacidade: tais sequências, bastante comuns nas línguas de referência, parecem corresponder a um mecanismo lexical que consiste em denotar referentes específicos por meio de significantes compactos.

Concretamente, a adjacência entre dois nomes numa relação de modificação identifica um referente mais específico, baseado em relações de hiponímia/hiperonímia, caracterizando, assim, a forma como o léxico reflete a especificidade de algumas áreas do saber³³. Assim as publicações científicas em inglês contêm, como salienta Fabre (1996), um número crescente de “*compound nouns*”, sendo muitos deles terminológicos (“*noun phrase, light year, city transportation, earth life, milk disease*”...). Além destes glossários, tem sido, no caso das línguas Românicas, nas redes sociais e na imprensa que esta tendência se tem desenvolvido com maior visibilidade. A partir de casos N1N2 extraídos dos corpora CETEM e FrWaC, vamos propor uma distinção entre “expressões binomiais compostas”, largamente maioritárias nestes corpora, e “expressões binomiais modificadas”, sequências comparativamente raras que são o objeto específico desta análise contrastiva³⁴.

2.1. Expressões binomiais compostas

Expressões binomiais compostas (EBC) implicam que, para além de N1, N2 é também um nome substantivo com capacidade referencial e, portanto, que se trata de casos de composição na perspetiva inicialmente descrita. Tanto em Português como em Francês, as EBC são interseccionais, como se ilustra em (7.): nestes casos, a adjacência entre os dois N equivale, por intersecção, a uma nova expressão referencial, de interpretação composicional, a considerar no plano semântico como um hipónimo de N1 e de N2.

³³ Cruse (2006) define da seguinte forma a noção de especificidade: “the property which distinguishes a hyponym from a hyperonym: the hyponym is more specific, the hyperonym more general. The hyponym gives more detailed information and denotes a narrower category. Thus, dog is more specific than animal, scarlet than red, and sprint than run” (2006, p. 167).

³⁴ Não atribuiremos nenhum valor específico à ortografia francesa ou portuguesa, já que, dada a novidade e origem dos dados, a mesma é bastante inconsistente.

(5.) a. EBC em Português (fonte: CETEMPúblico)

trabalhador-estudante, saco-cama, ministro-presidente, cidade-dormitório, cardeal-patriarca, actor-encenador, emissor-receptor, sobrinho-neto, juiz-conselheiro, artesão-empresário, casa-museu, cirurgião-dentista

b. EBC em Francês (fonte: FrWaC)

achat vente, bar tabac, camarade cheminot, chambres studios, citoyens membres, foire exposition, député-maire, ministre-président, chirurgien-dentiste

Resultando de uma situação de intersecção denotativa, os casos em (7.) devem aceitar diferentes testes morfológicos (como o plural em (8.)), mas recusar testes sintáticos como a inversão N2N1 em (9.):

(6.) EBC interseccionais: flexão plural

- a. trabalhadores-estudantes, juízes-conselheiros
- b. chirurgiens-dentistes, auteurs-compositeurs

(7.) EBC interseccionais: inversão N2N1

- a. cão pastor / *? pastor cão
- b. ministre-président / *? président-ministre

2.2. Expressões binomiais modificadas

Uma relação de modificação pode, mais raramente, juntar N1N2, resultando, neste caso, em expressões binominais modificadas (EBM). As mesmas implicam, portanto, que N2 é não-referencial e que estamos perante casos de modificação adjetival na perspetiva anteriormente descrita. As EBM aparentam, de facto, basear-se numa relação nome-adjetivo, ao contrário das EBC, e, como tal, devem ser distinguidas dos casos de denotação interseccional ou de composição, como os apresentados em (7.). Observemos em (10.) listas alargadas de EBM, retiradas dos *corpora* referidos³⁵:

(8.) a. EBM em Português – N1N2 (fonte: CETEM)

projecto-piloto, medida-padrão, clube-satélite, zona-tampão, retrato-robot, linha mestra, zona sul, país irmão, andar modelo, zona centro, cinema paixão, défice recorde,

³⁵ Ortografia original.

tecnologia laser, império fantasma, emissor pirata, código fonte, negócio-base, valor-guia, desporto-rei, região-piloto, frase-chave, ano recorde, rali-maratona, prova vedeta, astro-rei, cidade-berço, cidade martírio, factor novidade, homem orquestra, cidade património, fator risco, grupo piloto, programa candidato, torre montanha, músico anfitrião, efeito porto, organização fantecho, data limite, fator tempo, papéis vedetas, circuito satélite, pai herói, linhas mestras, elementos chave, questões base, lado lâmina, discos piratas, equipa maravilha.

b. EBM em Francês – N1N2 (fonte: FrWaC)

valeur seuil, taux limites, solutions internet, soirée spectacle, soirée cinéma, projets phares, puissance-crête, patrons voyous, partenariat pionnier, paradoxe clef, organisations sœurs, organes clés, objets tendance, niveau bac, niveau lycée, modules noyau, monde cauchemar, facteur temps, facteurs clés, exemple type, entreprises clientes, épouse modèle, effet plumeau, effet prix, effets lumière, effets troupeaux, écart type, démarche qualité, critères limites, roman fleuve, espace jeune, code source, candidat mystère, cellules souches, mot valise, chalutiers pirates, cheminées fantaisie, clause passerelle, bilan langage, bijoux fantaisie, avantages nature, astronomie passion, association relais, année charnière, âme sœur, albums cultes, affluence record, activités nature, actions carbone.

3. Modificação intensional

Do ponto de vista semântico, as EBM baseiam-se em processos denotativos situados, como vimos, nos limites da norma gramatical, em que N2 deriva frequentemente de um processo metafórico ou metonímico (como “*phare*” em “*measure-phare*” ou “estufa” em “efeito estufa”), para qualificar o hiperónimo N1³⁶. Nestas expressões, N2 passa por um processo de “atenuação”, que isola ou extrai um dos seus traços semânticos e o habilita a modificar N1³⁷. Assim, em “*measure-phare*” (lit. “medida farol”), o N2 “*phare*”, despojado de

³⁶ Nesse sentido, N2 é “específico” do hipónimo. É necessário conferir Cruse (2006) sobre as relações entre semântica lexical e especificidade.

³⁷ Jarema (2006, p. 47) fala de “transparência semântica” nos casos em que, num dado contexto, determinadas representações conceptuais de N2 são (des)activadas no composto, levando, assim, a distinguir entre compostos “transparentes” (em que N2 dispõe de todas as suas representações) e compostos “opacos” (em que apenas determinada representação semântica está disponível). Nas EBM, a relação de modificação é por defeito opaca.

valor referencial, destaca um dos atributos específicos do referente de N1, a sua visibilidade ou a sua ‘centralidade’. Da mesma forma, na expressão referencial “*fille canon*” (“uma brasa”, lit. “rapariga canhão”), a extensão de “*fille*” é especificada pelo N2 “*canon*”, que, privado de extensão, atua como modificador metafórico de N1. A possível ausência do modificador N2 não afeta, por isso, a denotação do hiperónimo N1: um N1N2 é um N1 – “um programa modelo é um programa”.

De forma a reforçar o grau de transparência³⁸ sugerido acima, a dicotomia ‘fregiana’ *extensão / intensão*, retomada por Carnap (1947) pode ser também evocada³⁹. Segundo Carnap “we take as the extension of a predicator the class of those individuals to which it applies and, as its intension, the property which it expresses” e acrescenta que “the extension of an individual expression is the individual to which it refers; its intension is a concept of a new kind expressed by it, which we call an individual concept” (1947, p. 1). Da mesma forma, Paiva Raposo (2013) refere, relativamente à oposição *intensão/extensão*:

“O significado de uma palavra (ou mais geralmente de uma expressão linguística) pode ser decomposto em pelo menos duas dimensões diferentes. Por um lado, um conjunto de propriedades que definem o conceito expresso pela palavra; por outro lado, a classe de entidades do mundo que satisfazem essas propriedades e para a qual remete o conceito expresso pela palavra” (Paiva Raposo, 2013, p. 187).

Assim, a semântica referencial de N1N2 implica consenso extensional sobre a “classe de entidades do mundo” identificadas por N1 e consenso intensional sobre “o conjunto de propriedades que definem” N1. Na interpretação de uma EBM, N1 é, de facto, em termos extensionais e intensionais um termo portador de referência e sentido, mas N2 remete apenas para a atribuição ao referente de propriedades consensuais. Nesse sentido, N2 pode ser interpretado com um *estereótipo*, cuja *imagem mental* se reduz a atributos relacionais e alegóricos – eventualmente apreciativos ou depreciativos. Como veremos no ponto seguinte, com base no glossário N2 identificado nos *corpora*, a interpretação de N2 nas EBM é intensional, isto é, nos termos de Paiva Raposo (2013), reduz-se a “propriedades que definem o conceito”, tendo N2 perdido a dimensão referencial relativa a “*entidades do mundo*”, dando assim legitimidade

³⁸ “Less lexicalization means more transparency, more lexicalization more opacity. More transparency implies more motivation of the compound via its members” (Dressler, 2006, p. 40).

³⁹ Frege (1892). Sobre propostas de aplicação à descrição gramatical, cf. Paiva Raposo (2013).

a uma leitura figurada (Paiva Raposo, 2013, p. 187). O uso de *modificadores intensionais*⁴⁰ N2 resulta estruturalmente de uma situação de adjacência superficial entre N1 e N2, em que N2 modifica a intensão de N1, dando assim origem a uma expressão derivada original. As EBM correspondem, pois, a expressões denotativas em que N2 funciona como modificador e N1 como expressão referencial. Alguns N1N2, como “*effet tunnel*” ou “*palavra-chave*”, adquirem estabilidade suficiente para se tornarem novas entidades lexicais, ao lado de compostos já consagrados como “*chien-loup*” ou “cão lobo”.

A geração de EBM é na sua maioria fortuita, ou seja, estas estruturas são forjadas de acordo com as necessidades inerentes a uma determinada enunciação: as EBM correspondem a formas lexicalmente frágeis, que proliferam provavelmente com base na transferência dos “*compound nouns*”, mas causando (ao contrário do Inglês) algum impacto pela substituição inesperada de um adjetivo por um nome modificador. Por serem expressões instáveis e acidentais, verificam uma interpretação não-composicional, baseada numa relação de modificação metafórica improvisada. Claramente, não resultam, ao contrário das EBC, de uma operação de justaposição ou concatenação. Uma vez institucionalizada, a EBM torna-se um item lexical familiar e recorrente, disseminado em vários tipos de contextos e usos, como por exemplo a linguagem administrativa, técnica, social, jornalística, etc., sendo que todos estes domínios manifestam especial atração pela compacidade, elegância e simplicidade destas estruturas lexicais inovadoras (“*démarche qualité, niveau bac, remède miracle* – fator tempo, conflito-fantasma, desporto-rei, país irmão”). Nestas estruturas modificadas, N1 é claramente o constituinte central, uma vez que atribui as suas propriedades semânticas, sintáticas e morfológicas relevantes a toda a estrutura N1N2 ou, por outras palavras, N1N2 herda as propriedades básicas (incluindo a referencialidade) do núcleo N1. Assim, um “conflito-fantasma” é um tipo de “conflito”, com o qual partilha as características sintáticas e semânticas (é um substantivo de género masculino), enquanto o N2 “fantasma” é um modificador que atribui a N1 uma interpretação marcadamente figurada.

Resultando de estruturas de modificação, as EBM correspondem assim, do ponto de vista formal, à projeção de um SN de núcleo lexical inicial N1, e resultam de princípios estruturais em que N2 um adjunto pós-nominal.

⁴⁰ “In modificative or attributive type the constituents are not linked by a dependency relationship: the non-head modifies and delimits the intension of the head or specifies an attribute of the head’s body. The parts of the compound are linked by a modifier-modified or a modified-modifier relation” (Rio-Torto, 2013, p. 23).

A descrição destes N1N2 parte, portanto, do pressuposto de que são estruturas sintáticas derivadas com base nas categorias lexicais N e A. Uma vez que N2 não é um adjetivo canónico, mas sim um “nome adjetivado”, podemos considerar que carece de categorização específica e poderá ser temporariamente referido como um modificador intensional⁴¹. Por outro lado, os *corpora* revelam que estes termos constituem, nas línguas referidas, um *glossário* fechado que convém contrastar.

4. Análise contrastiva do glossário N2

Do ponto de vista dos *corpora*, conclui-se que um determinado conjunto de *modificadores intensionais* é selecionado sistematicamente como N2 nas EBM. Vejamos em (12.) um glossário de N2 representativos do português e do francês, respetivamente:

(9.) a. N2 em Português (*fonte: CETEM*)

base, modelo, tipo, padrão, limite, mestre, piloto, chave, líder, pirata, fantasma, relâmpago, mãe, satélite, vedeta, maratona, maravilha, fantoche...

b. N2 em Francês (*fonte: FrWaC*)

modèle, type, limite, maître, pilote, clé, leader, pirate, fantôme, éclair, mère, satellite, vedette, record, marathon, merveille, source, relais, plafond, plancher, homme, femme, maison, robot, culte, frère, sœur, souche, source...

4.1. Traços léxico-semânticos de N2

Embora seja manifesta a relativa falta de diversidade lexical do glossário bilingue apresentado em (9.), é possível identificar, mesmo assim, determinados traços léxico-semânticos comuns às duas línguas. Assim, no caso do francês,

⁴¹ Claramente, N2 não subscreve alguns dos principais critérios formais relativos à definição da categoria adjetival. A título de exemplo, N2 não aceita projeções funcionais adjetivais como QP (Quantifier Phrase) ou DegP (Degree Phrase):

* “pai muito herói” / * “soirée très cinema”

* déficit mais recorde” / * “monde plus cauchemar »

Além disso, como vimos, N2 aceita marginalmente a categoria funcional AgrAP (*Agreement Phrase*). Sobre as projeções funcionais internas do Sintagma Adjetival e a sintaxe do adjetivo em geral (Martinho, 2007, p. 357).

alguns N2 tendem para uma função de natureza télica (envolvendo uma finalidade) que consiste em verificar em N1 a existência de propriedades, qualidades ou modelos (“*modèle, type, source, souche, mère...*”), eventualmente com base em processos figurados ou analógicos (“*pirate, maison, frère, robot, fantôme...*”). Contudo, N2 desempenha também uma função de natureza quantificadora que consiste em associar N1 a avaliação superlativa – ou até leitura hiperbólica (“*vedette, clé, maître, limite, marathon ...*”) ou a graus de medição não discreta (“*plafond, base, plancher ...*”). Da mesma forma, em português, encontramos N2 relacionando N1 com modelos e normas (“*modelo, padrão, tipo*”) e com aproximações analógicas baseadas em metáforas (“*pirata, robô*”), mas também N2 associando N1 a termos de comparação superlativos e a unidades de medição (“*chave, mestre, base, satélite, limite*”). Propomos que o glossário N2 seja, por isso, dividido em dois tipos semanticamente complementares: os *N2 relacionais* são ancorados em processos normativos de similitude e analogia entre N1 e N2 (ex: “cinema paixão, organização fantoche, equipa maravilha, – *bijoux fantaisie, monde cauchemar, astronomie passion*”). Os *N2 superlativos*, por seu lado, procedem de relações de medição e quantificação de N1 por N2 (ex: “linha mestra, prova vedeta, elementos-chave – *taux limites, écart type, organes clés*”).

A distinção anterior entre *N2 relacionais* e *N2 superlativos* permite estabelecer alguns paralelos pertinentes. Assim, embora não seja generalizável, é fácil verificar a existência, no léxico, de casos de adjetivos cognatos derivados dos *N2 relacionais*, como se ilustra em (13.). Inversamente, salvo raras exceções, não se verifica a existência no léxico de cognatos dos *N2 superlativos*, como é o caso de “*vedette, clé, maître, limite, marathon – mestre, chave, limite, recorde.*”

(10.) Cognatos de *N2 relacional*

a. *mère / maternel, fantôme / fantomatique, cauchemar / cauchemardesque*

b. *base / básico, património / patrimonial, herói / heroico, paixão / apaixonante*

A dissimetria anterior levanta duas questões distintas: (i) como se explica que, quando existe, o cognato adjetival não seja diretamente selecionado por N1? (ii) como justificar em termos de adjetivos cognatos, a dissimetria lexical entre *N2 relacionais* e *N2 superlativos*? No primeiro caso, a preferência acentuada das EBM por um N2 em detrimento de um cognato, como por exemplo no par “*maison mère*” / “*maison maternelle*”, leva-nos a concluir que N2 desempenha funções que um modificador adjetival não poderia desempenhar, sendo provavelmente a principal o facto de o *N2 relacional*, na sua qualidade de nome, permitir estabelecer com N1 relações de natureza metafórica ou analógica, fora

de alcance de um adjetivo. Assim, em “*maison mère*” (lit. “*casa mãe*”), “*mère*” predica uma relação de filiação específica entre o referente de N1 e N2, ao passo que, em “*maison maternelle*”, essa relação ficaria afetada pelo conjunto de significados e conotações associados ao adjetivo “*maternelle*”⁴². Podemos concluir, com base em Rio-Torto (2013), que são os traços “*predicativos*” de N2 que acabam por reforçar a coesão global e a singularidade das EBM⁴³.

Relativamente à dissimetria lexical referida (não existem cognatos de N2 *superlativos*), tal se deve provavelmente ao facto de estes N2 equivalerem a termos de quantificação, sendo os mesmos, como se disse, baseados em relações de comparação ou avaliação superlativa de N1 por N2. Nesse contexto, os N2 (“*vedette, clé, maître, limite, marathon – mestre, vedeta, chave, limite, recorde...*”) são interpretados como unidades de referência e equivalem a expressões de medição.

4.2. Análise estatística de N2

Os N2 em (9.) são, previsivelmente, de frequência muito desigual. Nos exemplos (11.a.b.) a seguir, o glossário N2 em Português e Francês é ordenado com base em frequências superiores a 10 concordâncias, por ordem decrescente. Além da frequência total das concordâncias N2, estas tabelas indicam também a frequência do lema plural de N2 e a percentagem de ocorrências totais⁴⁴:

⁴² Como mostram exemplos como “*plat maison*” e “*comida caseira*”, este paralelo não é, contudo, constante de uma língua para a outra.

⁴³ Rio-Torto (2013, p. 32) associa N2 explicitamente ao estatuto de “*predicador*”. Sobre o paralelo entre N2 e adjetivos relacionais cognatos (Martinho, 2013).

⁴⁴ A percentagem é relativa a N1 no plural exclusivamente. Assim, por exemplo, o N2 português “*base*” aparece no plural em 6 casos num total de 614 N1 no plural; da mesma forma, o N2 francês “*client*” aparece com o lema plural 742 vezes num total de 1599 N1 no plural.

(11.) a. Frequência (>10) dos N₂ em Português (fonte: CETEM)

N2	Freq. total	Freq. plural	Freq. total	Freq. plural	
base	1514	6	mãe	249	5
recorde	1315	55	relâmpago	156	2
limite	1108	45	vedeta	108	38
tipo	1108	0	padrão	105	1
chave	626	26	mestre	73	4
líder	452	65	fantochê	72	15
piloto	425	23	síntese	59	0
fantasma	376	82	maratona	26	0
modelo	365	10	maravilha	13	0
satélite	331	79	TOTAL	8777	613 (7%)
pirata	296	157			

(11.) b. Frequência (>10) dos N₂ em Francês (fonte: FrWaC)

N2	Freq. total	Freq. plural	Freq. total	Freq. plural	
clé	2381	6	maitre	118	28
client	1999	742	fantôme	105	21
type	1528	83	plafond	97	27
limite	1323	0	éclair	94	6
source	527	96	plancher	91	20
record	434	1	pirate	86	31
clef ⁴⁵	385	0	vedette	69	14
mère	383	57	roi	66	5
candidat	275	117	passion	64	2
centre	2676	20	robot	45	10
pilote	254	130	canon	35	10
modèle	211	21	reine	28	1
culte	199	56	martyr	23	6
leader	195	0	fleuve	18	1
satellite	161	9	marathon	13	0
lumière	124	25	TOTAL	11716	1558 (13,5%)

⁴⁵ Variante gráfica de “clé”.

Os exemplos seguintes, por seu lado, fornecem dados específicos sobre a frequência comparada de N2 nas duas línguas com base nos cinco N2 mais frequentes em (12.), e os cinco N2 mais raros em (13.):

(12.) N2 frequentes

Francês – 8135 (62%)	Português – 5671 (65%)
clé	base
client	recorde
type	limite
limite	tipo
source	chave

(13.) N2 raros

Francês – 117 (1%)	Português – 243 (3%)
canon	mestre
reine	fantoché
martyr	síntese
fleuve	maratona
marathon	maravilha

De notar, em primeiro lugar, que os cinco N2 mais frequentes representam 62% de todos os casos em francês e 65% em português, pelo que constituem o glossário essencial dos modificadores intensionais nas EBM. Quanto aos cinco menos frequentes, estão presentes em, respetivamente, 1% e 3% das concordâncias, o que os torna, na prática, irrelevantes e pontuais neste glossário. Do ponto de vista semântico, os exemplos anteriores mostram que os modificadores intensionais mais frequentes são, nas duas línguas, os N2 relacionais, ou seja, os N2 relativos a propriedades, modelos, regras e padrões consensuais. Os menos frequentes, por seu lado, correspondem aos *N2 superlativos*, isto é, a unidades de avaliação e ênfase (“*fleuve, reine, canon, marathon/maratona, mestre*”) e termos de comparação superlativos (“*clé/chave*”). Dominam, pois, neste glossário, os termos de interpretação relacional, cujo valor normativo e classificador é diretamente solicitado quando funcionam como N2: a ontologia de N2 incide predominantemente sobre normas e modelos (“*clé, type / chave*”) e acessoriamente sobre avaliação quantitativa específica (“*marathon/maratona, mestre*”). Uma conclusão óbvia deste contraste é que, em ambas as línguas, as EBM se alimentam em prioridade de *N2 relacionais*, os quais representam, como se observou, dois terços das concordâncias totais.

4.3. Análise morfológica

Do ponto de vista das lematizações, nota-se, no caso do francês, uma relação óbvia entre os *N2 relacionais* e a distribuição do lema plural. Assim, os *N2* com plural mais frequentes são “*clients, pilotes, candidats, sources, types*” (1168 concordâncias), termos relacionados com tipologia de géneros e propriedades. Inversamente, a frequência do lema plural para os *N2 superlativos* é baixa (“*reines, fleuves, satellites, passions, rois*” – 18 concordâncias) ou até nula (“*marathon*”) em francês. No caso do português, contudo, os *N2* com lema plural mais frequentes são precisamente os *N2 superlativos*, como “*recordes, limites líderes*” (493 concordâncias). Inversamente, a frequência do lema plural para os *N2 relacionais* é baixa (“*bases, mães, padrões*” – 12 concordâncias) ou até nula (“*tipo*”) em português. A distribuição anterior é resumida⁴⁶ em (14.):

(14.) Distribuição do lema plural por tipos de *N2*

	N2 francês		N2 português	
	relacional	superlativo	relacional	superlativo
N1 plural	<i>plural</i>	<i>singular</i>	<i>singular</i>	<i>plural</i>

Em resumo, não existe, em termos contrastivos, nenhuma diferença semântica significativa relativamente ao glossário *N2* incluído nas EBM extraídas dos corpora de trabalho. Na verdade, pode facilmente concluir-se que se trata de um recurso expressivo perfeitamente produtivo nas duas línguas, quer com *N2 relacionais* quer com *N2 superlativos*. Existe, contudo, uma diferença morfológica apreciável relativamente ao glossário *N2*, pois o francês tende a pluralizar os *N2 relacionais* ao passo que o português pluraliza em prioridade os *N2 superlativos*. Claramente, os *N2* parecem dissociar nas duas línguas o tipo semântico e a variação morfológica. Outra das conclusões a extrair do contraste destes dados é que a frequência do lema plural é globalmente baixa e desigual – é duas vezes mais frequente em Francês (13,5%). De facto, apenas 7% de *N2* aparecem no plural em português, o que, de alguma forma, confirma a observação de Villalva (2003), segundo a qual os “*modificadores nominais*” não flexionam nesta língua⁴⁷.

⁴⁶ Existem, no entanto, neste quadro, exceções significativas, em que *N2 relacionais* são de frequência elevada, mas de plural raro ou nulo. É o caso de “*clé*” e “*limite*” em francês, e de “*base*” e “*tipo*” em português.

⁴⁷ Sobre a importância e o valor dos padrões de concordância entre *N1* e *N2* (Rio-Torto, 2013).

5. Síntese

Em suma, é possível sugerir que, devido à sua reinterpretação relacional ou superlativa, os N2 parecem ter perdido a sua capacidade referencial, agindo claramente como modificadores intensionais de N1. Não se trata de entidades referenciais que, por intersecção, complementam a denotação de N1 (caso dos compostos), mas de um processo em que N2 manifesta propriedades reprodutíveis de modificador. É claro que, se a existência de N2 como “*pirate, limite, fantôme, type / recorde, chave, piloto...*” (com propriedades modificadoras indiscutíveis) não sofre contestação, por outro lado, o facto de existirem no mesmo conjunto (neste caso nos *corpora*) exemplos de N2 como “*fantaisie, fleuve, nature, tendance / património, rádio, montanha, martírio...*”, parece mais surpreendente. No entanto, estes últimos aparentam obedecer ao mesmo princípio de transparência (ou opacidade) composicional: perderam a sua capacidade referencial e atuam como modificadores, atribuindo ao referente de N1 (hiperónimo) uma leitura específica. Assim, poder-se-ia parafrasear, por exemplo, a EBM “*roman fleuve*” (lit. “romance rio” – *romance interminável*) como: “*um romance tão extenso quanto um rio*”, em que “*fleuve*” é, pois, um N2 *superlativo*. Da mesma forma, a EBM “*cidade martírio*” pode ser parafraseada como “*uma cidade que sofre as dores de um mártir*”, em que o N2 “*martírio*” seria *relacional*.

Por outro lado, pode prever-se que o processo aqui analisado não está de forma alguma esgotado no glossário identificado: seria possível, ao contrário daquilo que acontece com compostos, produzir novas EBM, resultando em combinações abertas, como as que destacamos⁴⁸ a seguir em (15.):

(15.) EBM – N1N2 não atestados nos *corpora*

- a. clube modelo, livro fonte, condenação exemplo
- b. bataille phare, livre clé, discours continent

A possibilidade de combinar livremente N1 com uma gama reduzida de N2 determina assim uma classe aberta de EBM, cuja notação ortográfica algo flexível reforça a relativa autonomia dos elementos combinados. Ao contrário dos compostos binominais inicialmente expostos em (7.), os casos em (10.) – também as EBM virtuais em (15.) – correspondem a estruturas de modificação

⁴⁸ Da mesma forma, é pertinente observar que eventuais EBM mais recentes - ou até novas - serão facilmente identificadas e processadas por qualquer falante nativo, o que confirma tratar-se de um processo criativo potencialmente capaz de sobregeração de combinações em número elevado.

originais, marginalmente lexicalizadas. Estas observações reforçam a nossa proposta de considerar N2 como o equivalente de um adjetivo. Em termos de estrutura interna do SN, N2 ocupa a posição adnominal típica de um modificador adjetival, ou seja, uma posição adjunta interna⁴⁹.

6. Conclusão

A existência de sequências EBM do tipo “equipa maravilha” ou “*candidat mystère*”, em que N2 é um objeto lexical e sintático ambíguo, levanta questões sobre a forma como tal item lexical seria categorizado. As EBM põem efetivamente em causa a informação lexical segundo a qual N2 é um item [+N] [-V], ou seja, uma categoria por natureza incompatível com a função de modificador. Como vimos, N2 manifesta, contudo, propriedades estruturais e semânticas que correspondem aos modificadores. A existência deste tipo de itens híbridos, meio-substantivos meio-adjetivos, representa certamente um desafio para a gramática e sugere a necessidade de repensar as categorias lexicais. Em todo o caso, tentámos demonstrar, numa perspetiva contrastiva, que algumas sequências endocêntricas N1N2 em Português e Francês são na verdade estruturas modificadas, onde N2 é um “*nome adjetivo*”, como ilustram “modelo” e “*bélier*” em “uma criança modelo” ou “*une voiture bélier*”. Nas construções binominais modificadas, considerámos, com base no estudo contrastivo das concordâncias extraídas dos corpora *CETEM* e *FrWAC*, que N2 partilha, tanto em português como em francês, a sintaxe, a semântica e (acessoricamente) a morfologia das categorias adjetivais. Além disso, observámos que, neste tipo de relação de modificação baseada em processos metafóricos comuns às línguas românicas, os N2 equivalem a modificadores relacionais ou superlativos.

Referências bibliográficas

BARBAUD, P. (1971). L'ambiguïté structurale du composé binominal. *Cahiers de Linguistique*, 1, 71-88.

⁴⁹ O estudo das projeções funcionais de N1 e N2 nas EBM é uma área de investigação potencialmente promissora (Rio-Torto, 2013; Martinho, 2019).

- BISETTO, A. e SCALISE S. (2005). The classification of compounds. *Lingue e Linguaggio*, IV (2), 319-32.
- BOOIJ, G. (2005). *The Grammar of words. An Introduction to linguistic morphology*. Oxford University Press.
- BREKLE, H. (1984). Ad hoc compounds in contemporary German: Pragmatic-Semantic reflections. *DRLAV*, 31, 97-106.
- CARNAP, R. (1947). *Meaning and Necessity. A Study in Semantics and Modal Logic*. The University of Chicago Press.
- COMRIE, B. (2003). On Explaining Language Universals. In M. Tomasello. (ed.), *The New Psychology of Language* (vol. II, pp. 195-209). Mahwah, N.J.: Lawrence Erlbaum Associates.
- DRESSLER, W. (2006). Compound types. In G. Libben and G. Jarema (eds.), *The Representation and Processing of Compound Words* (pp. 23-41). Oxford: Oxford University Press.
- FABRE, C. (1996). *Interprétation automatique des séquences binominales en anglais et en français. Application à la recherche d'informations*. [Doctoral dissertation, Université de Rennes].
- FREGE, G. (1892). Sense and Reference. *The Philosophical Review*, 57(3), 209-230.
- GREVISSE, M. & GOOSSE, A. (1961). *Le Bon Usage*. Editions Duculot.
- GUEVARRA, E. & SCALISE, S. (2009). Searching for Universals in Compounding. In S. Scalise, E. Magni & A. Bisetto (eds.), *Universals of Language Today* (pp. 101-128). Amsterdam: Springer.
- JAREMA, G. (2006). Compound Representation and Processing: A cross-language perspective. In G. Libben & G. Jarema (eds.), *The Representation and Processing of Compound Words* (pp. 45-69). Oxford: Oxford University Press.
- LIBBEN, G. (2006). Why study compound processing? An overview of the issues. In G. Libben e G. Jarema (eds.), *The Representation and Processing of Compound Words* (pp. 1-21). Oxford: Oxford University Press.
- LIBBEN, G. & JAREMA, G. (2006). *The Representation and Processing of Compound Words*. Oxford University Press.
- LINGUATECA (2008). *CETEMPUBLICO*. URL: <http://www.linguateca.pt/cetempublico/>
- MARTINHO, F. (2007). *Sintaxe e semântica dos adjetivos graduáveis em português*. [Tese de doutoramento, Universidade de Aveiro].
- MARTINHO, F. (2013). Noms épithètes dans les expressions binominales. *LINGUÍSTICA, Revista de Estudos Linguísticos da Universidade do Porto*, vol 8, 39-67.

- MARTINHO, F. (2019). The internal syntax of adjectival quantification in romance, *LINGUÍSTICA, Revista de Estudos Linguísticos da Universidade do Porto*. Vol. 14, 69-98.
- MONTERMINI, F. (2008). *La Composition en italien dans un cadre de morphologie lexématique*. Artois Presses Université, 161-187.
- NOAILLY, M. (1990). *Le substantif épithète*. PUF.
- PAIVA RAPOSO, E., BACELAR DO NASCIMENTO, M.F., MOTA, M.A.C., SEGURA, L. e MENDES, A. (2013). *Gramática do Português* (vol. I), Fundação Calouste Gulbenkian.
- RIO-TORTO, G.M. (1998). *Morfologia Derivacional. Teoria e Aplicação ao português*. Porto Editora.
- RIO-TORTO, G.M. (2013). Nouns in apposition: Portuguese data. *LINGUÍSTICA, Revista de Estudos Linguísticos da Universidade do Porto*. Vol. 8, 17-38.
- ROHRER, C. (1967). *Die Wortzusammensetzung in modernen französisch* [Doctoral dissertation, Universität Tübingen].
- VELOSO, J. & MARTINS, A.M. (2011), Etimologia não é morfologia: produtividade e composicionalidade na formação e processamento dos “compostos morfológicos” do português. In *Atas do XXVI Encontro Nacional da Associação Portuguesa de Linguística* (pp. 558-573. Porto: ENAPEL
- VILLALVA, A. (2003). Formação de palavras: composição, In M. H. Mateus, A.-M. Brito, I. Duarte, S. Frota, G. Matos, F. Oliveira, M. Vigário e A. Villalva. (2003). *Gramática da Língua Portuguesa* (pp. 969-983). Lisboa: Caminho.
- VILLALVA, A. (2008). *Morfologia do Português*. Universidade Aberta.
- VILLOING, F. (2012). French compounds. *Probus*, 24 (1), 29-60.
- UNIVERSITÀ DI BOLOGNA. (2008). *The CoLiTec corpora. FrWac Small*. URL: <https://bellatrix.sslmit.unibo.it/noske/public/#dashboard?corpname=frwac1>

TÍTULO: Análise contrastiva das expressões binominais em corpora anotados portugueses e franceses

RESUMO: Os adjetivos são modificadores atributivos e integram o sintagma nominal (SN) como adjuntos do núcleo N, compartilhando com este, nas línguas Românicas, as suas características flexionais. Alguns modelos de composição, no entanto, admitem um SN com dois nomes adjacentes, numa configuração que aparenta também ser de modificação. Partindo de propostas anteriores (Noailly 1990, Villalva 2003, Dressler 2006), a nossa análise implica que as sequências endocêntricas N1N2 em português e francês são, na verdade, estruturas modificadas, onde N2 corresponde a um “nome atributivo” (Martinho 2007, 2013). De fato, se uma sequência aleatória como * “voiture âge” é claramente inaceitável, expressões atestadas como “uma criança modelo” ou “une voiture bélier” parecem, pelo contrário, bastante claras sobre a natureza modificadora de “modelo” e “bélier”,

respetivamente, levantando contudo questões sobre a categoria, morfologia, sintaxe e semântica de tal modificador. Nestas “construções binominais modificadas” N1N2, frequentemente confundidas como compostos lexicais ou morfossintáticos, consideraremos, com base num estudo contrastivo das concordâncias extraídas dos corpora CETEM e FrWAC, que N2 parece partilhar tanto em português como em francês a sintaxe e a morfologia de uma categoria adjetival. Além disso, uma vez que N2 verifica propriedades alegóricas, mas carece de propriedades extensionais (Carnap, 1947), concluímos que, neste tipo de relação de modificação estereotípica, transversal a todas as línguas Românicas (Fabre 1996), os modificadores N2 de facto correspondem a adjetivos intensionais com interpretação relacional ou quantificada.

TITLE: Contrastive analysis of compound expressions in Portuguese and French annotated corpora
 ABSTRACT: Adjectives are attributive modifiers, and they integrate the Noun Phrase as adjuncts of a nominal head N, sharing with it, in Romance, inflectional features. Some compounding models, however, allow an NP with two adjacent nouns N1N2, in which seems to be an apparent modifying relation. Departing from previous proposals (Noailly 1990, Villalva 2003, Dressler 2006), our analysis implies that endocentric N1N2 sequences in Portuguese and French are in fact modified structures, where N2 is close to an “attributive noun” (Martinho 2007, 2013). Indeed, if a random sequence like * “voiture âge” would be clearly unacceptable, attested expressions like “uma criança modelo” or “une voiture bélier” seem, on the contrary, quite clear about the modifying nature of “modelo” and “bélier” respectively, however raising questions about the category, morphology, syntax and semantics of such a modifier. In these “binominal modified constructions”, frequently classified as lexical or morphosyntactic compounds, we will consider, based on a contrastive study of the concordances extracted from the CETEM and FrWAC corpora, that N2 appears to share in both Portuguese and French the syntax and morphology of the adjectival category. Moreover, since N2 meets allegorical properties but lacks extensional ones (Carnap, 1947), we conclude that, in this kind of stereotypical modification relation, shared by all Romance languages (Fabre 1996), N2 modifiers in fact revert to intensional adjectives with relational or quantified interpretation.

TITRE: Analyse contrastive des expressions binominales dans des corpus annotés portugais et français
 RÉSUMÉ: Les adjectifs sont des modificateurs attributifs, et intègrent à ce titre le syntagme nominal en tant qu’adjoints d’une tête nominale N, partageant avec elle ses caractéristiques flexionnelles. Certains modèles de composition, cependant, admettent un SN contenant deux noms adjacents N1N2, ce qui semble correspondre à une relation de modification. S’écarter de propositions précédentes (Noailly 1990, Villalva 2003, Dressler 2006), notre analyse implique que certaines séquences endocentriques N1N2 en Portugais et en Français sont en fait des structures modifiées, où N2 est proche d’un « nom attributif » (Martinho 2007, 2013). En effet, si une séquence aléatoire comme * “voiture âge” est clairement inacceptable, des expressions attestées comme “uma criança modelo” ou “une voiture bélier” semblent, au contraire, assez claires quant au caractère modificateur de “modelo” et “bélier” respectivement, soulevant cependant des questions sur la catégorie, la morphologie, la syntaxe et la sémantique d’un tel modificateur. Dans ces “constructions binominales modifiées” de type N1N2, souvent classées comme des compositions lexicales ou morphosyntaxiques, nous considérerons, sur la base de l’étude contrastive de concordances extraites des corpus CETEM et FrWAC, que N2 semble partager aussi bien en Portugais qu’en Français la syntaxe et la morphologie d’une catégorie adjectivale. De plus, puisque N2 vérifie des propriétés allégoriques mais est dépourvu de valeur extensionnelle (Carnap, 1947), nous concluons que, dans ce genre de relation de modification stéréotypée, partagée par toutes les langues romanes (Fabre 1996), les modificateurs N2 correspondent en fait à des adjectifs intensionnels pourvus d’une interprétation relationnelle ou quantifiée.