

CHILD-DIRECTED PUN-BASED COMMUNICATION IN CHILDREN'S PICTURE BOOKS – CASE STUDY

INTRODUCTION

The present study attempts a Cognitive Linguistics analysis of creative puns employed in the picture book for toddlers “I love you like no OTTER.” (Rossner, 2020). As the title suggests, the intended humorous effects of the play on words hinge upon recognition of the phonetic similarity between an animal word, i.e., *otter*, and another word, i.e., *other*, in a well-known phrase conventionally used in a confession of love, e.g., *I love you like no other*.

The focus of this analysis ties into rich research on language play, especially punning, contributed by linguists of various theoretical persuasions (e.g., Attardo, 1994, 2014; Lundmark, 2003; Partington, 2009; Zije, 2024). Our goal, in part, is to broaden the scope of existing analyses to include the Cognitive Linguistics paradigm with its recognition of cognitive operations as motivating factors that shape and guide meaning construction and interpretation (e.g., Croft & Cruse, 2004; Panther & Radden, 2011; Radden & Panther, 2004). Specifically, we claim that meaning processing in pun-based child-directed communication as found in picture books similar to the one under investigation involves figuration and is motivated, among others, by the cognitive operations of metonymy, metaphor and, especially, conceptual blending.

The framework adopted for this investigation is primarily Fauconnier and Turner's (2002) model of conceptual blending, yet the analysis is situated in broader research into and subsequent adaptations of their model (e.g., Brône & Coulson, 2010; Coulson & Oakley, 2005; Croft & Cruse, 2004; Feyaerts & Brône, 2005; Omazić, 2005). It particularly incorporates Brandt and Brandt's (2005) modifications.

Finally, in picture-book communication, meaning is communicated via two modalities: verbal and visual. The two complement each other and the resulting content communicates more than either can individually (Nodelman, 1991). In this analysis, it is demonstrated that both modalities contribute to setting up dynamically emerging complex networks of mental spaces in the operation of conceptual blending.

THE DATA

As mentioned, this is a case study of a small selection of texts and their illustrations in one picture book for toddlers. The book is rather short to meet the attention span of a zero- to three-year-old child. It is composed of two-page spreads with the text on the left-hand side and its illustration on the right. The texts are one or two sentences long, and each picture depicts an animal referenced in the text, often clearly a mother and baby of the same species. It also includes hearts, and the animals are engaged in some shared activity. The theme of the texts is the parent's love for the child. Each text is constructed on wordplay-based humor and the pleasurable experience of joined reading hinges on the child's recognition of the pun and its resolution.

As already mentioned, in this analysis, the texts and the visual materials are treated on a par – all contributing meaning in separate modalities. As convincingly argued by Nodelman (1991), this doubleness in communicating intended messages is built into the constructional framework of children's picture books. As a rule, illustrations do more than provide superfluous

support to the textual import by communicating different information (1991, p. 1). This is, in part, due to the fact that they rarely present the perspective of the autodiegetic narrator, in contrast to the narration conveyed by the text whereby “the same person is both the teller of the story and a key figure within it” (1991, p. 2). As a result, picture books convey yet another story, which is “the combination of the first two” (1991, p. 2). This composition of a child’s picture book is partly attributable to the fact that children experience hearing the words and seeing the picture as packages of information that are quite distinct from each other and that require alternating attention while processing them separately (1991, p. 1).

Finally, let us briefly discuss the pragmatic and situational context in which the studied material is used. Picture books are typically read in a familiar and friendly environment. The relative imbalance of authority is not invoked and the interactive and interpersonal effect that the parent aims to achieve is the strengthening of the parent-child relationship. This is accomplished via both the message of parental love conveyed by each text and the bond created by the shared pleasure derived from indulging in language play.

KEY METHODOLOGICAL ASSUMPTIONS

This study endorses as its key assumption the Cognitive Linguistics tenet that meaning is a cognitive (or mental) phenomenon. Thus, rather than a fixed and static semantic value residing in a linguistic unit, linguistic meaning is viewed as a conceptualization evoked by the unit in speakers’ minds, which is contextually construed in the ongoing discourse in every communicative act (Croft & Cruse, 2004, p. 3, 97; Langacker, 2013, p. 30). Dynamic and emergent, such conceptualizations encompass both established conceptions and novel ones, in “apprehension of the physical, linguistic, social, and cultural context” (Langacker 2013, p. 30). Furthermore, a linguistic unit (as a whole or its form or content [Radden & Panther, 2004, p. 23]) is believed to be motivated to the extent that some of its properties are shaped, or guided, by a variety of factors, including cognitive ones (Radden & Panther, 2004, pp. 24–31). The latter encompass metonymy, metaphor, and conceptual integration, in addition to non-figurative inferencing (Panther & Radden, 2011, p. 10). Conceptual metaphor and metonymy are defined respectively as a cognitive mapping (or a set of correspondences) across discrete conceptual domains based on analogy and correlation in experience; and a domain-internal mapping whereby one domain serves as a point of access to another within one overarching domain or domain complex, based on conceptual contiguity or association (Bierwiazzonek, 2013, p. 15, 30; Panther & Radden, 2011, p. 2). Particularly pertinent to the present analysis is conceptual integration (or conceptual blending), an on-line dynamic process whereby two or more input mental spaces, defined as “small conceptual packets constructed as we think and talk, for purposes of local understanding and action” (Fauconnier & Turner, 2002, p. 40), are projected to the output space, i.e., the blend, which develops its own ‘emergent’ content that is not supplied by the input spaces (Fauconnier & Turner, 2002, p. 42). According to its creators, the framework of conceptual blending explains, to a certain extent, how meaning construction and comprehension are, by nature, creative and imaginative.

CONCEPTUAL BLENDING THEORY

In the model advanced by Fauconnier and Turner (2002), conceptual integration involves at minimum two partially structured *input spaces*, whose matching elements are put in correspondence by *counterpart connections*. The generic space captures the structuring that appears to be shared by the input spaces and is then mapped onto those input spaces so that a “given element in the generic space maps onto paired counterparts in the two input spaces.” (2002, p. 47). The structure from the input spaces is *selectively projected* into the *blend*. Due to the processes of *composition* of the elements projected from the inputs, *completion* is

achieved by recruiting some background knowledge independently from the inputs, and, finally, *elaboration*, i.e., running the blend imaginatively, *emergent meaning* is generated in the blended space (2002, pp. 46–48).

The basic integration network can be modified to create a rich variety of blends, with four main kinds of conventionally distinguished networks: simplex, mirror, single-scope and double-scope (Fauconnier & Turner, 2002, pp. 120–135). The simplex network integrates (part of) a frame with roles determined by that frame, which partially structure one input space, and values supplied by the other input space. For illustration, in the process of conceptual integration, the frame *the family*, with its roles *father*, *daughter*, etc., is *compressed* with values such as *Paul* and *Sally*, creating the conception *Paul is the father of Sally*. The compression is facilitated by a Role-Value cross-space connection, which is one of several vital relations guiding compression in blending (2002, p. 92, 122). In a mirror network, all input spaces share a broad topology, i.e., an organizing frame. For example, all four spaces may be furnished with the frame: *boat sailing along an ocean course* (2002, p. 123). Thus, counterpart connections among spaces are easily established and compressions over vital relations are easily performed. However, at the level of finer detail, more specific frames may differ; e.g., a *nineteenth-century clipper on a freight run* in one input space and a *late-twentieth-century exotic catamaran on a speed run* in the second. In the blended space, the shared broad organizing frame may inhere in another frame, as a result of which, e.g., boats from different centuries may *race* against one another, thus producing rich emergent meaning that can be run imaginatively. A single scope network comprises input spaces that are each organized by different frames; however, only one of the frames is projected into the blend (2002, pp. 126-131). To demonstrate, in an utterance describing a business transaction between competing companies: *Murdoch knocks out Iacocca*, one input space is organized by the *business* frame and the other by the *boxing* frame, and the frame topology in the blend is inherited from the latter. The privileged frame has “prebuilt superb compression” (2002, p. 129) as we have general knowledge of the basics of *boxing*, the roles of boxers and the relations between them, as well as the causality of their dynamic interactions, etc., and the projections from the other frame are values fitting this frame without disrupting it. As a result, we acquire an insightful understanding of the otherwise diffuse knowledge of the *business* input space as the blended space remains in active connection with the entire network. A double-scope network integrates inputs organized by distinct frames, both of which contribute to the organizing frame of the blended space. If the topologies of the input spaces differ sharply, integrating select aspects may create rich clashes, resulting in highly creative blends. For example, the computer desktop interface, in which we *move* files to the *trash* and *save* documents into *folders*, is the result of the conceptual integration of aspects of two frames (inputs): *office work* and *computer work*. Part of “the imaginative achievement” in the formation of creative blends derives from juxtaposing clashing yet compatible frames in the inputs (2002, p. 131).

Fauconnier and Turner’s (2002) model of conceptual blending has been augmented by scholars working with the theory to better accommodate some aspects of meaning which, as has been argued, it falls short of explaining (e.g., Brandt & Brandt, 2005; Coulson & Oakley, 2005; Author, 2020). Some criticism is that the original model fails to clearly account for pragmatic meanings. For illustration, let us consider the oft-cited example of a blend in (1):

(1) This surgeon is a butcher.

In this integration network, the two input spaces are furnished with the frames of, respectively, *surgery* and *butchery*, with the counterpart correspondences couched in the generic space as: *agent*, *undergoer*, *sharp instrument*, *work space*, and the *procedure of cutting flesh* (Croft & Cruse, 2004, p. 208). The cross-space mapping between the input spaces puts in correspondence the respective parts from the frame of surgery: a surgeon performing in an operating theatre, using a scalpel to heal the patient; and the frame of butchery: a butcher

cutting the flesh of a dead animal with a knife in a butcher's shop in order to produce edible parts. The elements that are selectively projected to the blend and compressed create a scene whereby a butcher performs surgery on a human patient by using their butchery skills (2004, p. 209). It remains unclear specifically how the inference of incompetence and the pragmatic value of the utterance being a reproach or an insult arises, given that the conventional knowledge furnishing both inputs refers to two skilled professions (Brandt & Brandt, 2005, p. 223).

The above analysis was augmented with some additional spaces by Brandt & Brandt (2005, pp. 224–226), bringing in better apprehension of the current discourse context as well as the expressive value of the utterance. Among others, they add a reference space (providing an element with specific reference: *this surgeon*) and a presentation space (providing an element with generic reference: *a butcher*), as well as an interpretation space (relevance space), which, in the blend in (1), is furnished with a force-dynamic relation between an agent who causes harm and a patient. Additionally, they posit the need for an autonomous ethical schema, providing valuation in terms of right and wrong. Projections to the blend from these spaces contribute to an emergent meaning not of the incompetence of the surgeon but of ethically irresponsible behavior and normative negative valuation.

The core of Brandt & Brandt's (2005) modifications lies in the belief that the meaning of an utterance, shared between the speaker and the addressee engaged in shared attention, is interpretable in the particular situation of use and outside of the actual use it has no intrinsic meaning (Brandt & Brandt, 2005, p. 220). The proposed model appears to insightfully explore the recalcitrant instance of the blend in (1). Arguably, it also provides a schematic and broadly applicable framework for unpacking *ad hoc* blends, such as those described in the present paper. Specifically, Brandt & Brandt postulate a network whereby communication (or the expressive act, signification), couched in the model as a space of *semiosis*, always takes place within a specific situation, whose relevant aspects as framed by the participants feed into the integration as a space of *situation*. Both of these spaces are encompassed by the space of the *pheno-world*, which pertains to the broad universal knowledge of the human phenomenological life-world. In other words, communication operates from inside the pheno-world and in a specific situation, which together determine the semiosis in a speech event. The three spaces, i.e. pheno-world, situation and semiosis, form the *semiotic base space* for the utterance. From this space, two additional spaces are set up, *reference space* and *presentation space*, which, broadly, pertain to, respectively, actuality and virtuality. As already mentioned, in the 'surgeon-as-butcher' blend, they contain as their relevant element, respectively, a specific surgeon and a generic butcher. The reference space and the presentation space fuse in the blended space, the 'surgeon-as-butcher' blend, which is a virtual representation that specifies an aspect of the actual representation (Brandt & Brandt, 2005, p. 238). Finally, there's a *relevance space* that specifies what is *situationally* relevant in a particular occurrence of communication (tailor-cut interpretation of a given utterance) and impacts the process of meaning construction and comprehension. The six-space model can be re-shaped to include any other relevant spaces, such as the ethical schema in the 'surgeon-as-butcher' blend. As Brandt and Cronquist (2022) demonstrate, this updated blending model has wide applicability. For example, it proves useful for analyzing the rich, communicative ensembles of visual, gestural, performative (theatrical), vocal, textual and musical features in live concerts. The utility of culturally and contextually sensitive models of blending, like Brandt & Brandt's above, or Coulson and Oakley's (2005) model, has also been argued for in the Author's analysis of memes (2020, pp.11-12). Relevant to this study, her analysis also advocates for treating entrenched idiomatic phrases as inputs in their own right, as suggested in Omazić (2005).

Pertinent to the present analysis, conceptual integration models (e.g., Fauconnier & Turner's model) can be modified to accommodate instances of conceptual integration whose

felicitous unpacking hinges on the understanding of puns. For illustration, consider the compelling newspaper headline *US slowdown punctures Michelin's profits* (Feyaerts and Brône 2005, pp. 24–28). The use of the verb *puncture* serves as a metonymic point of access to the target: *tires*, which, in the current discourse space, is identified as tires manufactured by Michelin. At the same time, the verb *puncture*, which can be glossed as ‘making a small hole, especially in a tire, thus letting air out of it, which is accompanied by a visible lowering of the body of the vehicle’. The phrase *puncture profits* uncovers the metaphorical cross-domain mapping, or correspondence, based on the relation of analogizing, between the downward movement of a vehicle’s body and a worsening financial condition of a business (tire manufacturer Michelin) via the metaphor LESS IS DOWN (2005, pp. 24–28). In their description, Feyaerts & Brône refer to the metonymic and metaphorical mappings as *double grounding* in (an) additional input space(s) in the modified framework. Both metonymic and metaphorical meanings of *puncture* are accessed and accessible due to the fact that the blended space remains in active connection with the entire network. Incidentally, Brône and Coulson (2010) establish that headlines that involve conceptual integration based on double grounding both require greater effort in cognitive processing and grant greater reward as they are more pleasurable for the reader.

LANGUAGE PLAY: PUNS

A pun will be defined as a rhetorical use of a word that, due to its phonetic realization, has two opposing senses and involves incongruence followed by resolution. It is either spontaneously or deliberately authored to achieve humorous effects in the audience (Attardo, 1994, pp. 127–128; Partington, 2009, pp. 1795–1796). Puns have been of long-standing interest among linguists of various theoretical persuasions, who have generated distinct, yet in many respects compatible, accounts of the disambiguation process as well as the source of humorous effects in puns, as briefly overviewed below.

Giora (1991), for example, suggests that the resolution of the pun depends on *the graded salience* principle, which claims that the more *conventional* the interpretation of the punning word, the more likely it is to be accessed first (as more salient). The other interpretation can only be activated under some conditions, for example, if the more salient meaning does not fit the context. On her account, conventionality is associated with familiarity, frequency of use, prototypicality, and popularity of meaning, or its activation by the preceding context.

Attardo (1994, 2014), in turn, couches the essence of verbal humor, including puns, in terms of *incongruity* between two distinct yet compatible interpretations (scripts) that cannot be true at the same time. The two meanings are “spuriously appropriate in light of the joke’s setup” (Attardo, 2014, p. 612), but they oppose each other in some specific ways. Once the receiver realizes that the communicative act violates the Gricean cooperative principle, according to which the speaker’s contribution is assumed to be helpful, accurate and relevant, they are forced to change to non-*bona-fide* mode of communication, understanding that humor is intended.

According to Partington, punning can be accounted for in light of Sinclair’s fundamental contention that communication is based on two modes of the organization of language (Sinclair, 1987 as cited in Partington, 2009, pp. 1798–1799). In the first mode, discourse is viewed as composed largely of semi-fixed, pre-constructed blocks whereby the speaker selects a word together with its *primed* linguistic behavior: its co-occurrence restrictions (collocations), its semantic preferences (semantic associations), its grammar-based selection restrictions (colligation) and its functions in the text (textual collocation). This kind of language set-up is referred to as the idiom, or phraseological mode, and it is the default organization of language in meaning creation and interpretation (Partington, 2009, p. 1803). In the other mode, the speaker is confronted with an open-ended choice of words, in the sense that they construe the

utterance by making a word-by-word selection, and each word is considered as an individual unit of meaning independent of other meanings (2009, p. 1803). Partington contends that the essence of punning lies precisely in the enforced switching from the default mode to the open-choice mode (2009, p. 1799). Oftentimes, puns enforce the switch (or: *relexicalization*) from the figurative to the literal; e.g., in language play, the idiom *sitting on a small fortune* would need to be reinterpreted as someone's physical location on a valuable object. Also, the phrase may be explicitly altered in order to challenge the receiver to reconstruct a conventional idiomatic collocation; e.g., a playful use of *industrial resolution* builds on the collocation *industrial revolution* (2009, p. 1809). The analysis of puns as modified phrasal units is also supported by approaches within Cognitive Linguistics (e.g., Naciscione 2020).

In this paper, we claim that punning involves cognitive (or mental) processing. We further argue that deliberately created puns in picture books for toddlers, such as the case at hand, can be analyzed as complex, multimodal, conceptual blends, including metaphorical and metonymic blends. Attempts at employing the framework of Conceptual Blending to explain verbal humor include, among others, Coulson (2001), Hougaard, Oakley, and Coulson (2022), Feyaerts and Brône (2005), Jabłońska-Hood (2020), Lundmark (2003), and Ziyé (2024), who argue that processing humorous incongruity involves blending selected elements from input mental spaces so that the emergent meaning in the blended space creates humorous tension not found in either of the component inputs. For example, Lundmark's analysis of verbal and visual advertising puns involving metaphorical expressions leads her to believe that disambiguation of the incongruous meaning consists in (partial) resolution whereby the non-metaphorical and metaphorical interpretations are set apart but continue to coexist, causing humorous effects. For illustration, an advertisement of cat food based on the slogan *Something for your cat and you to chew on* is analyzed as including an input space in which a cat eats cat food, and another in which a man considers information, as an expression of the conventional metaphor IDEAS ARE FOOD. The generic space is taken to include an agent performing an activity upon an object. In the blended space, we get a humorous scenario of a man who is eating cat food. Elaboration of the blend may increase the humorous effect (e.g., we may imagine the man as formally dressed and being served by a waiter). Lundmark finds a relation between the accessibility of a generic space and the degree of humor in puns, concluding that metaphor-based puns that activate a generic space generally exhibit a low degree of humor, which can, however, be elaborated by visuals.

Puns are most conventionally classified as either *perfect puns*, whereby the words activating different interpretations are identical in their phonetic realization, and near-puns, or *imperfect puns*, whereby the words are similar phonetics-wise but not identical (Attardo, 2014, p. 613; Partington, 2009, pp. 1796–1797). Research shows that the (im)perfectness of the pun does not affect the receiver's experience of pleasure (Attardo, 2014, p. 613). Exact puns can exploit homonymy (e.g., *club* as 'association of people' or 'heavy weapon'), or homophony (e.g., *leek* and *leak*) (Partington, 2009, p. 1803). In near puns, the receiver is confronted with an altered phonetic realization of a given unit, e.g., *with fronds like these, who needs anemones*, and is forced to reconstruct the original phonetic representation of the conventional text, *with friends like these, who needs enemies* (Partington, 2009, p. 1803).

Finally, according to Attardo (1994), it is still uncertain to what extent and at what age a full understanding of puns develops in children, especially since the ability to notice linguistic incongruity appears to precede children's ability to understand punning. He concludes: "it would seem that all puns have something beyond incongruity, i.e., an element of resolution, however small and marginal." (1994, p. 143). However, the market success of pun-based picture books targeted at toddlers, even if indirectly, suggests that puns can be felicitously resolved and appreciated quite early, and they may give a pleasurable experience to the child and, perhaps, the adult.

CHILD-DIRECTED COMMUNICATION AS CONCEPTUAL INTEGRATION NETWORKS

While we adopt Fauconnier and Turner's (2002) model for our analysis, we also agree with Brandt and Brandt's (2005) contentions that (i) mental representations involved in conceptual integration are set up within the semiotic base space, including the space of situation and the space of pheno-world; and (ii) it is essential that the knowledge of the particular discourse ecology is fed into the integration network. In our analysis, the mental spaces set up for the purpose of local understanding are taken to partially include such situationally relevant elements as the child, the parent, and the affection that the parent holds for the child. We presuppose an idealized immediate situational context in which parent and child are sitting close to each other, perhaps the child is on the parent's lap, in a familiar and friendly environment, indulging in the shared activity of reading/listening to the story and looking at the pictures together.

Furthermore, some core situational context is provided through the visual mode. In the illustrations on each two-page spread, the parent and the child see some typical animal behavior (flying, swimming, grooming, playing) that is carefully selected for the young audience and anthropomorphized (parent and child of the species exchanging loving looks, kissing, petting, hugging, etc.). As we shall demonstrate below, the illustrations independently provide a conceptual contribution to the network of mental input spaces, and, consequently, are vital for overall meaning construction and interpretation.

The recognition of the situationally relevant contextual elements referred to above comes from our knowledge of the world as we conceive of it. Part of that knowledge stems from a folk model of young children's learning processes and the benefits of parent-child, book-centered interaction. It is widely believed that picture-book interaction is not only a social and socializing activity but it also promotes learning and facilitates understanding. Lovable animal characters in picture books, cute and cuddly, are particularly effective in performing these social and cognitive functions. As noticed by parents, and confirmed by research, children readily identify themselves with animal characters (Spencer, 2020). Thus, the knowledge of the cognitive and social functions of animal picture books is relevant to the ecology of the studied discourse and needs to be fed into the conceptual network.

Finally, the dual meaning of the punning word and subsequent expectation of humorous effects are also relevant for the construction and interpretation of the messages communicated in the studied texts and, thus, fed into the network.

THE SPACE OF SEMIOSIS

We shall begin with a panoramic overview of the expressive acts that constitute our language data. The texts each incorporate a pun, wherein a conventional, frequently used, idiomatic collocation (in Partington's sense) cannot be cognitively processed in default mode as it contains a word fused with an animal word, e.g., *I love you BEARY much*. Each text is emblematic of the parent-child affection idiom, including familiar declarations such as *I love you very much; little sweetheart; nobody could compare; seal it with a kiss; we belong together; you are always in my heart; and you make me happy*.

Each punning word exploits the phonetic similarity/identity of two source words; e.g., note the similarity between the interrogative pronoun *who* and owl hooting calls in their conventional English phonetic transcription in the blended *WHOO loves you?* Also, onomatopoeic words denoting cries or sounds produced by specific animal species, e.g., *squeak*, or *purr*, can be played off, as in *My special little SQUEAK-heart* and *We make the PURRfect pair*. Other salient attributes of specific species with potential for punning include verbs of

movement, e.g., *hop*, as in *You make me very HOPPY*; and body parts, e.g., paws, as in *PAWsitively amazing*.

All the above animal words—denoting body parts, calls, sounds, and movements—express concepts that are associated (contiguous) with the concept of each respective animal. In the text, each of these concepts serves as a metonymic vehicle, or point of access, to the metonymic target, i.e., the whole integrated overarching conceptual domain of our rich knowledge about a given animal. Moreover, they seem to be cognitively appropriate vehicles, pertaining to physical (not mental) human experience with animals and, particularly, to perceptually salient attributes of animals (specifically, vision, audition, and motion).

Incidentally, we have checked the priming of *hop* and *paw* in the Corpus of Contemporary American English: COCA (Davies, 2008-) to see whether they are indeed optimal metonymic vehicles as reflected in conventional language use. In the book, *hop* is used as a metonymic link to access the readers’ knowledge about kangaroos. In COCA, the noun collocates of *hop* are as follows: *hip* (in hip-hop) –1; somewhat unexpectedly, *time* –2; the first animal word, *bunny* –3; further, less frequent, *rabbit* –17 and *rabbits* – 81; finally, less frequent still, *kangaroo* – 97 and *kangaroos* – 112. The ‘hoppy’ pun is disambiguated by the visual material, which foregrounds a joey in mama kangaroo’s pouch. *Paw*, in turn, appears to be a more felicitous choice of metonymic vehicle as reflected in conventional language use: *dog* is the second noun collocate, the third being *cat*. The visual shows dogs playing in a field of flowers.

Most of the studied puns play off the names of animals, as *My love for you is strong. I SEAL it with a kiss.* and *We BEElong together. This I know is true*, or the pun in the title of the book *I love you like no OTTER*. There are few exact puns, homophones, or homographs (SEAL/DEER, in *You are so DEER to me.*). The remaining are all near puns (e.g., SQUEAK/sweet, BEARY/very, BEE/ [be]long).

The above considerations provide the setting for an in-depth analysis of the specific instance *WHOO loves you?* below. We shall start with an analysis of the formal blend and proceed with a thorough description of the complex networks of mental spaces in the operation of conceptual blending.

WHOO LOVES YOU?

The question *WHOO loves you?* in the text in the first spread is a very salient blend formed from the conventional representation of the (reduplicated) hooting call of an owl: *hoo hoo /hu: hu:/* and a well-entrenched idiomatic collocation *who loves you?*. The parent would most probably imitate the hooting sound, perhaps even exaggerating it, while reading the question in order to bring out the incongruity and thus the pun. The phonetic similarity between one part of the owl’s call and the interrogative pronoun *who* facilitates the deliberate creation of a lexical (formal) substitution blend, as marked in Figure 1. Let us emphasize that the blend draws on the full collocation rather than an individual linguistic unit.

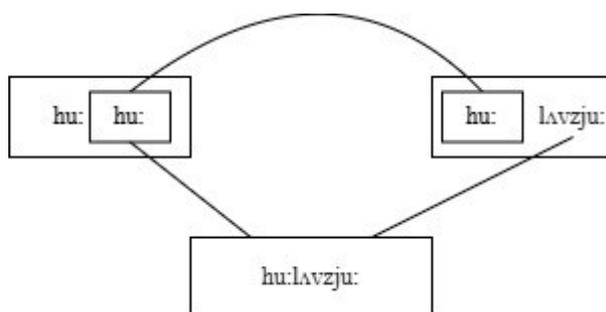


FIGURE 1. The lexical (formal) substitution blend of *WHOO*

At the conceptual level, prompted by both the language form and the visual material, the blend arises in a complex network of mental spaces as in Figure 2. We shall argue that it predominantly includes an input space furnished with a frame ‘owls’ (Input 1) and an input space ‘human family’ (Input 4). In Input 1, the hooting animal call—as conceptually associated with the animal producing it—serves as a metonymy-based point of access to the whole domain ‘owls’, containing general knowledge about owls, such as the owl’s nesting habits, mating, hatching, fledglings, etc. The input space is a partial and temporary construct furnished with elements of that general and stable knowledge. Apart from the calls, the input will especially contain the scenario of a parent raising an owlet and the roles of a *parent* owl and a *young* owlet. Arguably, the body of knowledge available for recruitment to Input 1 is generic in nature (in the sense of Brandt & Brandt’s [2005] presentation space). In the special circumstances of the interlocutors and their interpersonal relations, which we shall couch as part of Brandt & Brandt’s semiotic base space, we postulate the need for Input 2, set up by the visual clues whereby the anthropomorphic owls in the visual material (loving looks, physical proximity of the adult owl to the owlet, the owl preceding the young as if leading or protecting it) are put in correspondence with their counterpart roles from Input 1, as values. In the simplex integration network, they are connected, projected to the blend and compressed due to the vital relation of Identity, becoming a loving mama/daddy owl and their beloved baby owl from the picture in the toddler’s book (in the sense of Brandt & Brandt’s reference space), which, as we run the blend, are both flying in the night sky, exchanging loving looks and indulging in conversation (which is in animal code), confessing love, etc. The blended space is further recruited for integration as an input space (Input 3). Let us turn now to the ‘human family’ space (Input 4). By analogy with the ‘owl’ space, it is a presentation space, which is generic in nature, drawing from the feltwork of general and stable knowledge of the social construct ‘human family’, including roles such as *father*, *mother* and *child*, as well as the ability to communicate. Input 5, in turn, is a reference space. It provides the mental representation of a specific reader and a specific toddler-listener engaged in the pleasurable joint activity of reading a specific children’s book in the here and now of the speech situation. As Inputs 4 and 5 become integrated, the counterpart elements are connected and projected to the blended space, where they are compressed due to the vital relation of Identity (Input 6). In the blended space, they become parent and child engaged in the joint activity of reading and listening.

Before we bring together the ‘individuated owls’ space and the ‘individuated parent-child’ space, let us ponder the interpretation of the expressive act: *WHOO loves you? Let me tell you this*. In the interactional and interpersonal circumstances of the speech situation, the text is the opening of a story about the love that any parent feels for their offspring and how that love is reciprocated by the child, which unfolds on the successive pages. It is a confession of love that the parent makes using a communication code supplied by lovable animal characters, such as owls. We take the parent-child love relationship as well as the frame ANIMALS IN CHILDREN’S STORIES to be elements furnishing the relevance space (Input 7), which feeds into the blended space (Space 8). Also, the relevance space needs to include an appreciation of punning and the expectation of sharing a laugh.

Inputs 3 and 6 share a broad topology, i.e., organizing frame, which is also projected to the blend (Space 8). The integration network is then a mirror blend. Space 8 is essentially a virtual reality in which the reader and the young listener *are*, as if, a parent owl and a baby owlet, enjoying flying together, engaging in a conversation that is partly in human language and partly hooting. The parent owl is telling the story, and they obviously love each other. Importantly, they share the experience of recognizing the incongruence and resolving the pun.

Finally, we must discuss the generic space in our blend, which is absent from Brandt & Brandt’s framework. They argue it has no explanatory value as the shared structure is found in

the inputs (Brandt & Brandt, 2005, p. 246). In the mirror network, which we propose for the case at hand, the shared topology is particularly straightforward: it invokes two communicating animate entities remaining in a relationship ($e_1 \leftrightarrow e_2$). However, in *ad hoc* formal blends such as the above, the phonetic realization of the punning word, shared between the two sources (a cluster of sounds representing an animal call and a phonetic realization of the collocation), appears to have its own independent cognitive identity, beyond the shared context-sensitive structure of the component mental spaces. The awareness of the shared phonetic material (a similar cluster of sounds) and it being a punning word is brought to the fore in the cognitive processing of the incongruous text while the recognition of the shared organizing frame and analogy (perhaps identity for the child) comes secondary (more as a cause-and-effect nexus). Thus, we postulate the need for a generic structure which crucially contains the phonetic realization of the pun (Input 9), as represented in Figure 2.¹

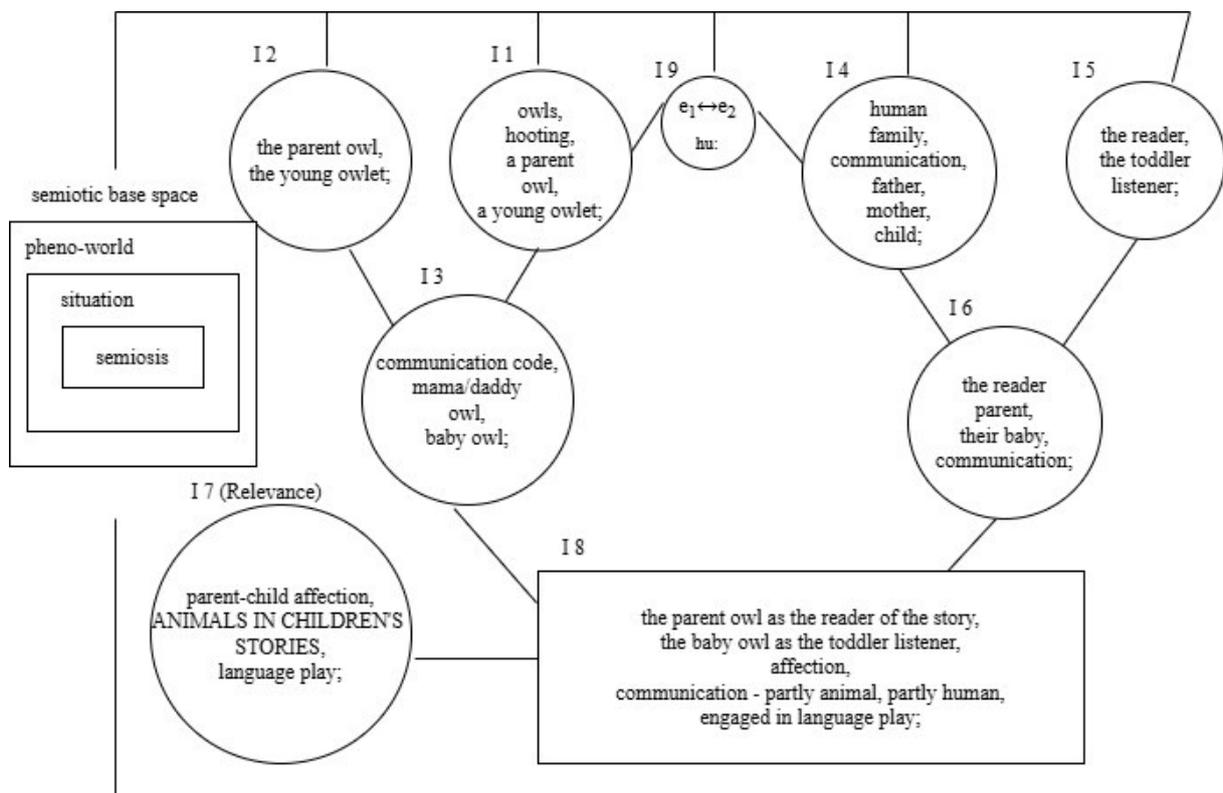


FIGURE 2. The complex network of conceptual blending in the pun *WHOO loves you?*

METAPHOR-MOTIVATED MEANING IN VERBAL AND VISUAL MODALITIES

So far, we have focused on a conceptual integration network that includes metonymy-driven compressions among its inputs. Feeding into such networks, there are also mental spaces that involve analogizing and compressing corresponding elements contributed by distant (i.e. discrete) domains, e.g., the physical domain of spatial distance (specifically, physical closeness) and the abstract domain of emotional intimacy. Thus, the conceptual integration networks underlying both the verbal and the visual data may also include metaphorical blending. In (2-4), we provide some examples of metaphor-driven verbal content (for clarity of presentation, we now avoid animal words), together with the respective metaphors.

¹ For clarity of presentation, we have simplified the scheme by leaving out the counterpart connections between spaces.

- (2) I miss you when we are apart. → EMOTIONAL INTIMACY IS PHYSICAL CLOSENESS (note also the simile *You stick to me like glue.*)
- (3) My love for you is strong. I *seal* it with a kiss → EMOTIONAL BONDING IS PHYSICAL BONDING
- (4) You are so *dear* to me. → LOVE IS A VALUABLE OBJECT.

The verbal metaphorical content is reinforced by the visuals, in each consecutive picture, the animal characters are either in direct physical contact (e.g., the joey in mama kangaroo's pouch) or remain close. The emotional-physical metaphorical relation is further reinforced in the real-time reading scenario where the child is likely cuddling with the parent and/or sitting on his/her lap, thereby embodying the metaphor in the book. Such physical instantiation of a metaphor has been shown to increase its effectiveness at conveying the intended meaning (e.g., Xu, Liu, & Wang, 2023; Khatin-Zadeh et al., 2023).

CONCLUSIONS

This study highlighted, among others, (i) the utility of the conceptual blending model in the analysis of punning; (ii) the importance of the situational context, immediate or otherwise, in humorous meaning creation and interpretation, and, thus, the utility of models which incorporate context as a core component; (iii) the relevance of the generic space in an analysis of pun-based creative formal blends. Let us comment briefly on these respective points.

When confronted with some incongruity of meaning, and, hence, the need to switch to a non-*bona-fide* mode of communication, the toddler-listener is challenged to resolve it by reconstructing the idiomatic phrase and setting it apart from the animal word itself. The humorous effect arises precisely from the interplay (or tension) between the two senses, both of which remain accessible. The framework of conceptual blending provides a model of cognitive processing whereby an emerging integrated network retains all interconnections among its spaces, including the blended spaces. As such, it is well suited for modelling the cognitive processing of two incongruent senses—both separately and in tandem— all accessible and accessed simultaneously. As we demonstrated, some counterpart connections and compressions in the network may be metonymic and/or metaphorical in nature.

Furthermore, outside the situational context of the speech situation and the broader folk knowledge of the cognitive and social functions of parent-child, animal picture book-centered interaction, the studied puns would be rather unimpressive and the humorous effects they produce rather weak. Yet, disambiguated by the visuals, and placed in the situational context, they become cute and funny, tailor-made for the special audience they are intended for.

Moreover, in our analysis of the blend *WHOO loves you?* as a mirror network, we highlighted the relevance of a generic space, claiming that the shared broad topology of the input spaces 'individuated owls' and 'individuated parent-child', which we described as two communicating animate entities in a relationship ($e_1 \leftrightarrow e_2$), is in the case at hand not as straightforward as typically found in mirror networks. Specifically, we claimed that the awareness of the shared structure appears to be secondary and backgrounded, while the shared phonetic structure of the formal blend is foregrounded and cognitively salient.

Finally, the paper contributes to research on a severely understudied discourse type, i.e., child-directed communication, especially that involving language play. Thus, it is hoped that the paper both identifies and begins to fill an important gap in linguistic research.